Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 585 801 A2**

EUROPEAN PATENT APPLICATION

(21) Application number: 93113602.2

(22) Date of filing: 25.08.93

(51) Int Cl⁵ **C12N 15/12**, C07K 13/00, G01N 33/68, C07K 15/28, A61K 39/395, G01N 33/577, A61K 37/02

(30) Priority: 28.08.92 JP 230028/92

(43) Date of publication of application:
09.03.94 Bulletin 94/10

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB IT LI LU NL PT SE

(71) Applicant: **HOECHST JAPAN LIMITED**
C.P.O. Box 1256
Tokyo 100-91(JP)

(72) Inventor: **Takeshita, Sunao**
1-40-12, Keyakidai
Tokorozawa-shi, Saitama(JP)
Inventor: **Okazaki, Makoto**
Mezon-Ishida 305,
3-12-42 Asahicho
Kawago(JP)
Inventor: **Kawai, Shinji**
1-12-27 Tsurma
Fujimi-shi, Saitama(JP)
Inventor: **Tsujimura, Atsushi**
105 Corpo Haruta,
5 Goshokaido
Mozume(JP)
Inventor: **Amann, Egon, Dr.**
5-1-13 Komazawa
Setagaya-ku, Tokyo(JP)

(74) Representative: **Losert, Wolfgang Dr. et al**
HOECHST AKTIENGESELLSCHAFT,
Zentrale Patentabteilung,
Gebäude F 821
D-65926 Frankfurt am Main (DE)

(54) Bone-related cadherin-like protein and process for its production.

(57) A bone-related protein named OSF-4 which is obtained from bone tissue of a mammal including mouse or human, and a process for its production. This protein is a novel naturally occurring mammal protein of the cadherin family.

OSF-4 acts as an adhesion molecule or a growth factor which takes part in the process of osteogenesis at the site of bone induction. OSF-4 can be used as an agent for treating bone metabolic diseases, and its high organ specificity for bones enables its use as a diagnostic reagent for bone metabolic diseases.

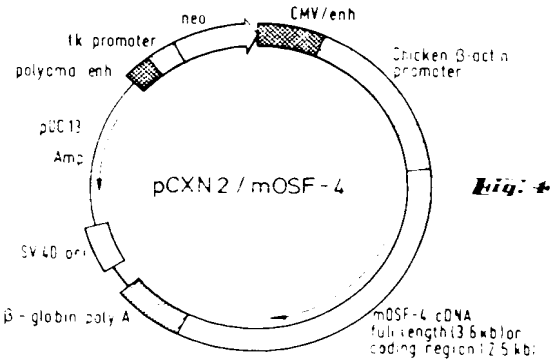


Fig. 4

EP 0 585 801 A2

This invention provides a novel bone-related protein. It is named OSF-4 and belongs to a group of cadherin molecules. The OSF-4 can be obtained from bone tissue of a mammal including mouse or human. This bone-related protein is useful for the diagnosis and treatment of bone metabolic diseases.

Bone metabolic diseases include osteoporosis, Paget's disease, osteomalacia, hyperostosis, and osteopetrosis. Osteoporosis, in particular, has a high incidence enough to affect about more than a half of postmenopausal women and elderly people, and effective methods for its diagnosis and treatment have been strongly desired.

Bone metabolic diseases involve some disorder of bone metabolism at the cellular level in bone tissue. The discovery, isolation and identification of factors associated specifically with bone metabolism are very effective for elucidating this disorder.

A cell line of osteoblasts which play a major role in osteogenesis, has been used and a proteinaceous factor produced specifically by this cell line has been identified. Therefore, the present invention provides a novel protein named OSF-4 which is substantially bone-specific, and which is highly homologous with various known cadherin type cell adhesion molecules in terms of amino acid sequence.

The OSF-4 can also be produced from the DNA sequence described in the present specification by an ordinary genetic engineering technique known in the art. Furthermore, the OSF 4 or its fragment can be produced from the amino acid sequence described in the specification by a chemical peptide synthesis method. Moreover, that fragment of the DNA sequence of the OSF-4 described in the present invention which is specifically different from other cadherin molecules can be synthesized with a length of 15 to 50 bases by an ordinary chemical oligonucleotide synthesis method. That fragmentary sequence can be used as a DNA probe for finding and identifying bone-derived cells. This identification of bone-derived cells is useful particularly for grasping the origin of metastatic or recurrent carcinoma, thus leading to an appropriate therapy for recurrent cancer. Of the partial peptides of the OSF-4, the peptide in the epitope portion that can be recognized by antibodies is usable for preparing a monoclonal antibody specific for the OSF-4. The resulting monoclonal antibody is of marked value for identifying bone-derived cells by an immunological cell tissue staining method. Because of its similarity to cell adhesion molecules, the OSF-4 is also useful for the treatment of fracture.

OSF-4 is a bone-specific cadherin-like proteinaceous factor. The following is known about cadherin which is a cell adhesion molecule involved in morphogenesis

The segmentation of cell population is one of the most basic elements for the construction of an animal body. This segmentation begins at a very early stage of morphogenesis. As the differentiation of cells proceeds, the same types of cells migrate and become reorganized in an orderly manner, thereby performing morphogenesis as well as the construction and maintenance of tissues. One of the elements that control such cellular migration is the selective adhesion of cells. Cells have the features of recognizing adjacent cells or adjacent extracellular matrices, and adhering to only particular ones. In accordance with the differentiation of cells, their adhesion specificities vary. Consequently, these cells may leave particular sites, migrating to and gathering at the sites where they should have originally been situated. So far, numerous cell adhesion molecules have been identified and all show cell type specificities. These molecules can be roughly classified into more than 4 groups, i.e. cadherin family, immunoglobulin superfamily adhesion molecules, integrin superfamily, selectins, and those not belonging to these categories (Hynes et al., (1992) Cell, vol. 68, pp. 303-322).

Cadherins, in particular, have been well analyzed, and not only their structures, but their functions have also been extensively studied. Cadherins are glycoproteins with molecular weights of about 120 kD, and are Ca^{2+}-dependent intercellular adhesion molecules (Takeichi, (1991), Science, vol. 251, pp. 1451-1455). Nearly 10 types of them have been identified. The respective types have binding specificities, and the same molecules react homophilically with each other. As a result, cells having the same type of cadherin bind selectively to each other. Thus, cadherins are considered indispensable for determining the specificities of intercellular adhesion. Typical examples are E-cadherin (epithelial cadherin), P-cadherin (placental cadherin), N-cadherin (neural cadherin), and L-CAM (liver cell adhesion molecule). All of them are similar in structure; each is composed of an extracellular region comprising 4 to 5 repeats of about 110 amino acids, a transmembrane region, and a cytoplasmic region comprising about 150 amino acids. Comparisons among the respective subclasses have shown about 50% identity of the amino acids. In the extracellular region, the identity rate is higher at a site nearer the N-terminus; it becomes gradually lower the nearer the transmembrane region; and it is maximal in the cytoplasmic region. In recent years, new cadherins have been reported, including, for instance, M-cadherin (muscle cadherin), B-cadherin (brain cadherin), T-cadherin (truncated cadherin), and desmoglein localized in desmosomes. These cadherins may originate from a single ancestor gene, and constitute the cadherin family.

Close studies of a cadherin molecule have shown that sites binding to calcium ions and sites determining binding specificity are present in its N-terminal region. The cytoplasmic domain has been found to be functionally important for the adhesion property, and to bind to a protein such as catenin or actin. Through these functions, cadherins have been suggested to contribute to cytoplasmic signaling. During invasion and metastasis of cancer cells, these cadherins undergo quantitative changes, and so their relationship with oncogenesis has attracted a broad attention. Despite many such reports of cadherin molecules, there have been no reports of cadherin molecules with specificity for osteoblasts.

Bone formation and maintenance are dependent on the balance between osteoblasts which form bones and osteoclasts which resorb bones. Osteoblasts are mesenchymal cells of the same origin as myoblasts and adipocytes while osteoclasts originate from stem cells as do neutrophils and macrophages. Osteoclasts are known to express vitronectin receptors which belong to the integrin family. With osteoblasts, the presence of cell adhesion molecules has only been suggested.

The object of the present invention is to find a new type of cell adhesion factor which is specifically expressed in osteoblasts. Such a new cell adhesion factor is an important molecule for the proliferation, differentiation, migration and reorganization of osteoblasts. This substance can be expected to find use in the diagnosis and treatment of various bone metabolic diseases.

cDNA of mouse OSF-4 (mOSF-4) was isolated from a mouse osteoblastic cell line MC3T3-E1 cDNA library constructed by a combination of PCR (polymerase chain reaction) and the subtraction method, and by differential screening. Then, the mouse OSF-4 cDNA was used as a probe for screening cDNA libraries obtained from human osteosarcoma cells. As a result, two types of OSF-4, named hOSF-4-1 and hOSF-4-2, were obtained, and their nucleotide sequences were determined. The nucleotide sequence of the OSF-4 is very well conserved between mouse and human. Comparisons between mOSF-4 and hOSF-4-1 as well as hOSF-4-2 show 97.1% and 96.4% identity in amino acid level, respectively (Tables 1 to 3). The very high conservation between these species suggests that OSF-4 has essential roles in vertebrate bone metabolism. OSF-4 can be isolated and purified from the bone extracts of other vertebrates. The OSF-4 of other animal species can be obtained from cDNA libraries or genomic libraries constructed from their bones, cultured bone cells and other body tissues by recombinant gene technology using the cDNA of the present invention or its DNA fragment as a probe. Search through the currently available DNA and amino acid sequence data bases demonstrated the sequence of the cDNA in the present invention to be novel.

In comparison with the amino acid sequence of mOSF-4, hOSF- 4-1 shows high conservation in the whole domain. hOSF-4-2, on the other hand, is completely identical with hOSF-4-1 in terms of the N-terminal to 631st amino acid residues. Because of the insertion of 179 bases into the transmembrane region, however, frameshift occurs, with the result that the 62 amino acid residues ranging from the 632nd amino acid residue to the 693rd-position C-terminus assume a completely different structure (Tables 1 to 3). Hence, the C-terminal 9 amino acid residues in the transmembrane region, and the cytoplasmic region are completely different between hOSF-4-1 and hOSF-4-2. Such cadherin with the C-terminal region deleted corresponds to T-cadherin. This type of cadherin has been suggested to take part in the control of cell adhesion.

A peptide corresponding to 15 hydrophilic amino acid residues at the 101st to 115th positions in the EC1 domain of mOSF-4 was chemically synthesized. This peptide was conjugated with KLH (keyhole limpet hemacyanin), and used for immunization of rabbits. The resulting anti-mOSF-4 peptide antisera were used for immunohistochemical detection of OSF-4 in systemic slices of the neonatal mouse. OSF-4 was detected in the osteoblasts, chondrocytes and so on.

Generally, the OSF-4 can be directly extracted from bone tissue or cartilage tissue of a human, bovine, murine or other source by a known biochemical technique.

The DNA coding for the OSF-4 can be obtained by constructing a cDNA library or a genomic library from mRNA extracted from vertebrate bone tissue, and using a probe comprising a labeled fragment of the mouse DNA sequence disclosed in the present specification. A full length cDNA clone can be obtained by a combination of the above-described and other standard techniques on molecular biology.

As described above, OSF-4 shows homology with known representative cadherin molecules, but it is a cadherin molecule belonging to a new subclass different from those so far reported. Its structure is composed of 5 repeats in an extracellular region, a transmembrane region, and a cytoplasmic region (Fig. 1). Comparisons between OSF-4 and other cadherins have shown that homology in the extracellular region becomes lower from the N- terminus toward the transmembrane region and the highest homology is noted in the cytoplasmic region, according to the homology pattern among the existing different cadherin molecules (Table 4).

Table 4

Comparisons of amino acids among mouse OSF-4 and other cadherin molecules								
Types of cadherin compared	Homology (%)							
	EC1	EC2	EC3	EC4	EC5	TM	CP	MP
OSF-4:N	39.3	46.8	32.8	29.8	27.4	59.4	50.3	39.4
OSF-4:E	32.7	40.4	30.7	30.8	17.2	34.4	47.1	33.9
OSF-4:P	34.6	35.8	36.0	29.1	19.0	23.1	43.8	33.3
OSF-4:M	31.8	38.5	27.6	26.0	25.4	34.4	45.1	33.2
N:E	60.2	53.5	45.1	43.9	27.0	46.9	64.1	49.8
N:P	51.9	51.8	45.1	49.5	27.9	30.8	57.7	47.3
E:P	65.7	61.1	52.7	46.7	38.9	53.8	79.7	58.7

In Table 4, the homology of amino acids in each region was calculated and expressed in %. The abbreviations are as follows: EC1 to EC5, five extracellular regions; TM, transmembrane region; CP, cytoplasmic region; MP, mature protein; N, N-cadherin; E, E-cadherin; P, P-cadherin; and M, M- cadherin. In the column "Types of cadherin compared," OSF-4:N denotes comparisons of the amino acid sequences in the respective regions between OSF-4 and N-cadherin (the same is true for the other combinations).

The protein provided by the present invention is a group of glycoproteins, named OSF-4, which belongs to a new cadherin subclass and plays an important role in osteogenesis. More concretely, the human and mouse OSF-4 proteins described in this specification are included. OSF-4 is expressed in osteoblasts during the process of bone formation, and acts as a cell adhesion molecule and a morphogenesis-related substance. These human and mouse OSF-4 proteins can be used to identify and isolate other mammalian OSF-4 proteins similar in DNA sequence and amino acid sequence.

The present invention further provides polypeptides comprising analogues of OSF-4, i.e. mutants and fused proteins having OSF-5 adinty, as well as fragments of the OSF-4 which can be identified as OSF-4 related, particularly with at least 10, preferably 15 amino acids. The cDNA of mouse OSF-4 isolated from the mouse osteoblastic cell line MC3T3-E1 encodes a protein consisting of 796 amino acids, including a signal peptide composed of 24 amino acid residues. There are two isoforms of human OSF-4 which were isolated from a human osteosarcoma cDNA library. One cDNA clone, human OSF-4-1, encodes a protein consisting of 796 amino acids including a signal peptide composed of 24 amino acid residues. The other cDNA clone, human OSF-4-2, encodes a protein consisting of 693 amino acids including a signal peptide composed of 24 amino acid residues. The present invention also provides a process for producing OSF-4 by recombinant DNA technology.

According to the present application the term "hybridization under stringent conditions meers hybridization conditions with a salt concentration of 6 x SSC (NaCl-citrate puffer) at 62-68 °C.

Brief Explanation of Tables and Figures

Table 1 shows an alignment of the amino acid sequences of mouse OSF-4, human OSF-4-1 and human OSF-4-2. Common amino acid residues are shown in the form of consensus.

Table 2 shows a continuation of an alignment of the amino acid sequences of mouse OSF-4, human OSF-4-1 and human OSF-4-2 shown in Table 1. Common amino acid residues are shown in the form of consensus.

Table 3 shows a continuation of an alignment of the amino acid sequences of mouse OSF-4, human OSF-4-1 and human OSF-4-2 shown in Table 2. Common amino acid residues are shown in the form of consensus.

Figure 1 is a schematic drawing of the structure of mouse OSF-4 precursor protein. OSF-4 precursor protein is divided into eight regions, a signal region (shaded part), five extracellular regions (EC1, EC2, EC3, EC4 and EC5), a transmembrane region (TM) and a cytoplasmic region (CP).

Figure 2 shows a restriction enzyme map of cDNA coding for mouse OSF-4. The bold letters indicate the region coding for the amino acid of OSF-4. There are no KpnI and SalI sites in the map.

Figure 3 shows the tissue-specific expression of mouse OSF-4. This was analyzed by purifying RNA from various tissue and cultured cells followed by RNA dot blotting. This diagram shows the results of autoradiography.

Figure 4 shows the map of expression vector pMSS60. It is mentioned that the content of the Japanese

priority application No. 230028192 is also part of the present application.

Examples

The present invention will be described in more detail by reference to the following Examples.

Example 1

Construction of cDNA library by subtraction and PCR

The construction of a cDNA library specific for the osteoblastic cell line MC3T3-E1 will be hereinafter described. This cDNA library is constructed from MC3T3-E1 cDNA library by a combination of the subtraction method and the PCR with the gene expressed in mouse liver tissue being subtracted. Each cDNA clone has gene fragments with an average length of about 300 bases, and is characterized in that the gene with a low content has been amplified too.

Unless otherwise specified, all general recombinant DNA protocols complied with Sambrook et al., "Molecular Cloning Manual" (1989), Cold Spring Harbor Laboratory, Cold Spring Harbor, U.S.A. Total RNAs were extracted from 8×10^7 MC3T3-E1 cells and about 1 g of mouse liver tissue by the guanidine method. Poly A$^+$RNAs were purified from the total RNAs by means of the commercially available product "Oligo dT Latex mRNA Purification Kit" (Takara Shuzo). cDNAs were synthesized by a cDNA synthesis kit (Amersham) using 1 μg of each poly A$^+$RNA as a template. However, a random primer was used, instead of an oligo dT primer, in an amount of 1.5 times its ordinary amount used, whereby the cDNA chain elongation was restricted to an average length of about 300 bases. After the cDNAs were made double-stranded and blunt-ended by use of the above kit, they were joined with T4DNA ligase (Takara Shuzo) to the following two DNA linkers, i.e. ATOS-1.2 (Sequence ID Nos. 4 and 5 of the Sequence Listing) for the MC3T3-E1 cDNA, and ATOS-4.5 (Sequence ID Nos. 6 and 7 of the Sequence Listing) for the liver cDNA:

ATOS-1/2:

ATOS-1 5′- CTCTTGCTTGAATTCGGACTA-3′

ATOS-2 3′-ACACGAGAACGAACTTAAGCCTGAT-5′

ATOS-4/5:

ATOS-4 5′- CTCTTGCTTAAGCTTGGACTA-3′

ATOS-5 3′-ACACGAGAACGAATTCGAACCTGAT-5′

Then, each reaction product was subjected to DNA amplification by the PCR (polymerase chain reaction) method using ATOS-1 and ATOS-4, respectively, as primers. The amplified DNA concentration was determined with the DNA assay kit "DNA Dipstick" (Invitrogen). The subtraction method was performed using photobiotin (Pirce). Photobiotin (20 ng) was added to 20 μg of the PCR-amplified liver cDNA, and light from a sunlamp 10 cm apart was projected onto the liver cDNA for 10 minutes to label it with biotin. To 3.0 μg of the labeled liver cDNA was added 0.3 μg of unlabeled MC3T3-E1 cDNA for hybridization. Then, streptavidin (Takara Shuzo) was reacted, and the reaction mixture was extracted with phenol to remove cDNA common to the liver cDNA from the MC3T3-E1 cDNA. The subtraction method was repeated to remove as much of the common cDNA as possible from the MC3T3-E1 cDNA. DNA was amplified by PCR using the aforementioned ATOS-1, and the DNA concentration was measured. This cDNA (10 ng) was digested with the restriction enzyme EcoRI, and then ligated with T4 ligase to 1 μg of the phage vector lambda gt10 (lambda gt10 EcoRI cloning kit, Stratagene) which was digested with EcoRI and dephosphorylated at its ends. The resulting recombinant DNA was packaged into lambda phage particles by use of the in vitro packaging kit "Gigapack-gold" (Stratagene). The recombinant phages were infected into E. coli C600 (preserved as HT003 at Japanese Cancer Research Resources Bank, National Institute of Health of Japan), and the organisms were applied to an agar medium along with a soft agar medium to form phage plaques. The efficiency of infection was determined to be 3×10^5 phage plaques/μg vector DNA.

The resulting cDNA library was subjected to differential screening to select clones with a high specificity for MC3T3-E1. Concretely, 2.25×10^4 phages were applied to total 10 plates, and the resulting

plaques on each plate were transferred to two nylon membrane filters (total 20 filters). These series of plaques were subjected to plaque hybridization with radiolabeled MC3T3-E1 cDNA as the probe for one of the series, and with radiolabeled liver cDNA for the other series. In 273 clones, expression was observed with the MC3T3-E1 cDNA probe, but not with the liver cDNA probe. These clones were used as a mini-library in subsequent experiments.

Example 2

Isolation of mouse OSF-4 clone

A description will be made of methods to identify a cDNA fragment of OSF-4 as an MC3T3-E1 specific clone from the mini-library constructed in Example 1, and to clone full length cDNA from the cDNA library of MC3T3-E1 with the use of this fragment.

The total RNAs from MC3T3-E1 and liver prepared in Example 1 were spotted in an amount of 1 µg each onto nylon membrane filters. 273 of the filters were prepared, and used for hybridization to be described later on. Separately, the DNA of the inserts of the 273 phage clones prepared in Example 1 was amplified by PCR. This DNA was agarose gel electrophoresed, and main bands were cut out, purified, and radiolabeled for use as a probe. A clone showing expression with MC3T3-E1 cDNA but no expression with liver cDNA upon autoradiography was recloned into a plasmid vector. Concretely, the DNA of the inserts amplified by PCR and then purified was digested with the restriction enzyme EcoRI, and recloned into the EcoRI site of the plasmid vector pUC118 (Takara Shuzo). The DNA sequence of the resulting clone was determined with commercially available "DNA Sequence Kit" (Takara Shuzo) using a universal primer. Search through DNA and protein data bases showed that DNA sequence to constitute a clone homologous with the existing cadherin. This clone was designated as D45, and used for subsequent cloning of the full length cDNA.

For cloning of the full length cDNA, blund-ended double- stranded cDNA was synthesized with the cDNA synthesis kit "cDNA Synthesis System Plus" (Amersham) using 5 µg of the poly A^+RNA of MC3T3-E1 purified in Example 1. The resulting cDNA was ligated to EcoRI/NotI adaptor (Takara Shuzo) using T4 ligase, and the product was agarose gel electrophoresed to purify a fragment more than about 700 base pair long. This fragment was joined to the EcoRI site of lambda gt10 phage vector (Stratagene), and packaged into phage particles in the same way as in Example 1. The packages were infected into E. coli as in Example 1, and the efficiency of infection was determined to be 1.5×10^7 phage plaques/µg vector DNA. The aforementioned D45 was radiolabeled for use as a probe, and 1.0×10^6 phage clones of the cDNA library were screened by plaque hybridization. Fourteen positive hybridization signals were obtained, whereafter the NotI fragment of the phage clone with the longest insert was recloned into the NotI site of the plasmid vector pGEM11Zf(+) (Stratagene). The resulting clone was designated as pKOT164.

Example 3

Determination of mouse OSF-4 DNA sequence

Deletion mutants of the pKOT164 and a subclone containing its cDNA fragment were prepared with "the Deletion Kit for Kilo Sequence" (Takara Shuzo) by cutting at intervals of about 300 base pairs in each opposite direction. The DNA sequence of each deletion mutant was determined with the automatic DNA sequencer 373A (Applied Biosystems, U.S.A.). The entire DNA sequence of the cDNA, and an amino acid sequence translated from this DNA sequence are shown as Sequence ID No. 1 of the Sequence Listing. The protein encoded by this cDNA was designated as OSF-4. No. 1 of the amino acid residue corresponds to the N-terminus of the predicted OSF-4 precursor protein. The restriction enzyme map of that cDNA is shown in Fig. 2.

Example 4

Tissue specific expression of mouse OSF-4

RNA dot blotting was performed to investigate the tissue specific expression of mouse OSF-4. The total RNAs of the thymus, spleen, brain, kidney, liver, lung, testis and heart of mice (purchased from Nippon Clea) were prepared by the guanidine method. Calvarial osteoblast-rich cells were obtained from a culture of newborn mice calvaria. Total RNA was extracted from these cells in the same way as described above.

One µg of the total RNA each from the above-mentioned tissues, cultured calvarial cells, MC3T3-E1 and mouse fibroblast cell line NIH3T3 (ATCC CRL 1658) was dotted onto nylon membrane filters (Biodyne, PALL) fixed by baking, and used for hybridization. Separately, the pKOT164 was digested with NotI, and isolated and purified by agarose gel electrophoresis. Then, the isolate was radiolabeled and used as a probe. Autoradiography indicated high expression for the cultured calvarial cells and MC3T3-E1, and low expression for the lung and testis (Fig. 3).

Example 5

Cloning of cDNA coding for human OSF-4

The NotI fragment containing the cDNA region of pKOT164 was purified and used as a probe to screen a human osteosarcoma cDNA library consisting of 1.3×10^{6} clones. Twenty-one positive signals were obtained, and 5 clones were isolated. Two clones with large inserts were recloned into plasmid vector pHSG398. The resulting plasmids were designated as pKOT161 and pKOT170.

Example 6

Determination of human OSF-4 DNA sequence

From the pKOT161 and pKOT170 cloned in Example 5 and their subclones, deletion mutants were prepared in the same way as in Example 3. Then, their DNA sequences were determined. These DNA sequences and the amino acid sequences predicted from them are shown in Sequence ID Nos. 2 and 3 of the Sequence Listing. The proteins encoded by these cDNAs were designated as human OSF-4-1 and human OSF-4-2. The amino acid residue No. 1 of each of them corresponds to the N-terminus of the predicted OSF-4 precursor protein.

Example 7

Preparation of anti-OSF-4 antisera

In preparing anti-peptide antibodies against mouse OSF-4, the corresponding 15 amino acid residues in the EC1 were synthesized by the solid phase synthesis method using a peptide synthesizer (430A, Applied Biosystems), in accordance with an experimental report on M-cadherin (Donalies et al., (1991), Proc. Natl. Acad. Sci., U.S.A., vol. 88, pp. 8024-8028). The synthetic peptide was OSF-4. 1 (FVIDDKSGNIHATKT, Sequence ID No. 8 of the Sequence Listing). This synthetic peptide was conjugated with KLH (keyhole limpet hemacyanin) using glutaraldehyde as a coupling agent, and used for immunization of rabbits. The resulting antisera could be used to search immunohistochemically for the presence of OSF-4 in newborn mouse systemic slices, and to detect the expression of OSF-4 in E. coli, yeast and animal cells.

Example 8

Expression of OSF-4 in animal cells

The present example describes the preparation and expression of the expression vector for mouse OSF-4 in animal cells and the functional analysis of the produced OSF-4.

There is an open-reading frame in the 5'-flanking region of the OSF-4-coding region in the base sequence as shown in SEQ ID NO: 1 in Sequence Listing. The open-reading frame was expected to decrease the translation efficiency of OSF-4. Therefore, a clone which contained the OSF-4-coding region alone was selected from the delation mutants prepared in Example 3, and it was used for the preparation of the expression vector for OSF-4. The segment of the clone from G of the 191st to A of the 2700th in SEQ ID NO: 1 in Sequence Listing was cut and a linker containing XhoI and BamHI sites and a linker containing XhoI site were ligated to the 5'-terminus and 3'-terminus of the segment, respectively. Then the segment bordered with XhoI site was inserted into the XhoI site of an expression plasmid vector for animal cells, pCXN2 (Niwa et al., (1991) Gene, vol. 108, p193-200). The OSF-4 expression vector obtained was termed pMSS60 (Fig. 4).

The pMSS60 was introduced into L-cells of a fibroblast cell line derived from mouse epidermis by the calcium-phosphoric acid co-precipitating method. Then 12 G418 resistant colonies transformed by pMSS60

were cultured separately to obtain the cloned cell lines. RNA was extracted from these cloned cells. Then, three OSF-4-high producer cloned cell lines were selected with an RNA dot blotting method by using mouse OSF-4 cDNA as the probe. The three clones were termed C1, C7 and C11, respectively.

A band of approximately 100 kDa reacting with anti-OSF-4 antibody was detected by Western blotting analysis out of the proteins produced in these cloned cells.

Furthermore, the functional analysis of OSF-4 produced by these cloned cell lines was conducted as follows by Takeuchi's aggregation assay method which had been originally established to examine cadherin cell adhesive properties (Takeichi et al., (1977) J. Cell Biol., vol. 75, p464-474).

First, from each monolayer cells (C1, C7 and C11), TC-and TE-treated cell suspensions were prepared.

The TC-treated cell suspensions were prepared as follows. The monolayer cells were rinsed three times with CMF solution (Puck's Ca^{2+} and Mg^{2+} free saline; J. Exp. Med. vol. 108, p954- 956, 1958) containing 1 mM calcium chloride. Then, the cell suspensions were incubated at 37 °C for 15 min with HCMF solution (HEPES-buffered saline; 8.0 g of NaCl, 0.4 g of KCl, 0.09 g of $Na_2HPO_4.7H_20$, 1 g of glucose, 2.38 g of HEPES, and 4.8 ml of 1N NaOH in 1,000 ml of H_20, pH7.4) containing 0.01% trypsin and 0.1 mM calcium chloride.

The TE-treated cell suspensions were prepared by the same procedure as for preparation of TC-treated cell suspensions except incubation at 37 °C for 15 min with HCMF solution containing 1 mM EDTA and 0.01% trypsin.

Then, each cell suspension was rinsed with CMF solution twice and divided into cell suspensions containing 1×10^6 cells/3 ml in HCMF solution with or without 1 mM calcium chloride. Each cell suspension was transferred to a 15 ml-size conical tube and was allowed to make cell aggregation by stirring with 80 rpm at 37 °C for one hour. The ratio of N1/N0 was calculated by counting with a Coulter-counter the number of cells before stirring (N0) and the number of cell clots after stirring (N1). The results are shown in Table 2; the OSF-4- expressing cloned cell lines showed calcim-dependent cell adhesive property similar to cadherin. Cadherin molecule is known not to be digested by trypsin with calcium. Thus, the cell suspension treated with EDTA (TE-treated) were digested by trypsin and did not show any cell adhesion.

[Table 3]

```
            601                                               650                                                    700
mOSF-4      LLSCNAEAYI LNAGLSTGAL IAILACIVIL LVIVVLFVTL RRQKKEPLIV FEEEDVRENI ITYDDEGGGE EDTEAFDIAT LQNPDGINGF IPRKDIKPEY
hOSF-4-1    LLSCNAEAYI LNAGLSTGAL IAILACIVIL LVIVVLFVTL RRQKKEPLIV FEEEDVRENI ITYDDEGGGE EDTEAFDIAT LQNPDGINGF IPRKDIKPEY
hOSF-4-2    LLSCNAEAYI LNAGLSTGAL IAILACIVIL LGCPSLMEPP SPREDMRLLY L......... .......... .......... .......... IPRKDIKPEY
Consensus   LLSCNAEAYI LNAGLSTGAL IAILACIVIL L----L---- -------L-- .......... .......... .......... .......... ........GF

            701                                               750                                                796
mOSF-4      QYMPRPGLRP APNSVDVDDF INTRIQEADN DPTAPPYDSI QIYGYEGRGS VAGSLSSLES ATTDSDLDYD YLQNWGPRFK KLADLYGSKD TFDDDS
hOSF-4-1    QYMPRPGLRP APNSVDVDDF INTRIQEADN DPTAPPYDSI QIYGYEGRGS VAGSLSSLES ATTDSDLDYD YLQNWGPRFK KLADLYGSKD TFDDDS
hOSF-4-2    QLMLFSYVKV NRRFCLLGVF IKLPFLYVVA TESPTTLTSL .......... .......... .......... .......... .......... ......
Consensus   Q-M------- ---------F I--------- --------S- .......... .......... .......... .......... .......... ......
```

SEQUENCE LISTING

General Information:

Applicant:

Hoechst Japan Limited
New Hoechst Building
10-16, Akasaka 8-chome
Minato-ku, Tokyo
107 Japan
Tel. (03) 3479-5137
Fax. (03) 3479-7859

Title of Invention: Bone-related Cadherin-like Protein and Process for its Production

Number of Sequences: 8

Computer Readable Form:

Medium Type: 3,5" HD Diskette

Computer: 386 SX

Operating System: MS-DOS

Software: ASCII

"Sequence Listing"

SEQ ID NO: 1
SEQUENCE TYPE: nucleic acid
SEQUENCE LENGTH: 3581 base pairs

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULAR TYPE: cDNA to mRNA

ORIGINAL SOURCE
ORGANISM: Mus musculus
STRAIN: osteoblastic cell line MC3T3E1

FEATURE:
SEQUENCE Description: SEQ ID NO: 1:

```
GAATTCGCGG CCGCCTTTGA AGACATTCAG TTCTGTTATT TATTGAATGA CCAATCAGAT      60

GGGTGGAGCA TGTTATAGGA ATTGGCAGCA GGTATCCAAT GGGTGAAGAA GAAGCTGACT     120

GCGGAGGTGA CCAACCCTGG CGTGATGTCC TCAGTGAGTG AAGATATTCC ATTCCAGAGG     180

AGGTCTACTT GACACATCTG GGAGGCCGCC ATCCGAAAGA AAGCCACTCT GTTGGTGTAG     240

GGAGTGACAG CTGCATTCTC CTGTGCCTAC TGCATAACCA AAA ATG AAG GAG AAC       295
                                                 Met Lys Glu Asn
                                                   1

TAC TGT TTA CAA GCT GCC CTG GTG TGC CTG AGC ATG CTA TAC CAC AGC       343
Tyr Cys Leu Gln Ala Ala Leu Val Cys Leu Ser Met Leu Tyr His Ser
  5                  10                  15                  20

CAG GCG TTT GCT CTG GAG CGA CGA AGC CAC CTG CAT CCC TCT TTC CAT      391
Gln Ala Phe Ala Leu Glu Arg Arg Ser His Leu His Pro Ser Phe His
                  25                  30                  35

GGA CAC CAT GAG AAG GGC AAG GAG GGG CAG GTG CTG CAA CGC TCC AAG      439
Gly His His Glu Lys Gly Lys Glu Gly Gln Val Leu Gln Arg Ser Lys
             40                  45                  50

AGA GGC TGG GTC TGG AAC CAA TTC TTT GTG ATA GAA GAG TAC ACC GGG      487
Arg Gly Trp Val Trp Asn Gln Phe Phe Val Ile Glu Glu Tyr Thr Gly
         55                  60                  65

CCT GAC CCT GTG CTG GTG GGC AGG CTT CAT TCT GAC ATT GAC TCC GGT      535
Pro Asp Pro Val Leu Val Gly Arg Leu His Ser Asp Ile Asp Ser Gly
     70                  75                  80

GAT GGG AAC ATT AAA TAC ATT CTC TCA GGT GAA GGA GCG GGA ACC ATT      583
Asp Gly Asn Ile Lys Tyr Ile Leu Ser Gly Glu Gly Ala Gly Thr Ile
 85                  90                  95                 100

TTT GTG ATT GAT GAC AAA TCA GGG AAC ATT CAT GCC ACC AAG ACA TTG      631
Phe Val Ile Asp Asp Lys Ser Gly Asn Ile His Ala Thr Lys Thr Leu
                105                 110                 115
```

```
GAC CGA GAG GAG AGA GCC CAG TAC ACA CTG ATG GCT CAG GCG GTG GAC      679
Asp Arg Glu Glu Arg Ala Gln Tyr Thr Leu Met Ala Gln Ala Val Asp
        120                 125                 130

AGG GAC ACC AAC AGA CCA CTG GAG CCA CCT TCA GAA TTC ATT GTT AAG      727
Arg Asp Thr Asn Arg Pro Leu Glu Pro Pro Ser Glu Phe Ile Val Lys
    135                 140                 145

GTC CAG GAC ATT AAT GAC AAC CCT CCA GAG TTT CTG CAT GAA ATC TAT      775
Val Gln Asp Ile Asn Asp Asn Pro Pro Glu Phe Leu His Glu Ile Tyr
150                 155                 160

CAT GCC AAT GTG CCT GAG AGG TCC AAT GTG GGA ACA TCA GTT ATC CAA      823
His Ala Asn Val Pro Glu Arg Ser Asn Val Gly Thr Ser Val Ile Gln
165                 170                 175                 180

GTG ACA GCC TCT GAT GCA GAT GAT CCC ACC TAT GGA AAT AGT GCC AAG      871
Val Thr Ala Ser Asp Ala Asp Asp Pro Thr Tyr Gly Asn Ser Ala Lys
                185                 190                 195

TTA GTG TAT AGC ATC CTT GAA GGA CAA CCC TAT TTC TCG GTG GAG GCC      919
Leu Val Tyr Ser Ile Leu Glu Gly Gln Pro Tyr Phe Ser Val Glu Ala
            200                 205                 210

CAA ACA GGT ATC ATC AGG ACA GCC CTT CCC AAT ATG GAC AGA GAA GCC      967
Gln Thr Gly Ile Ile Arg Thr Ala Leu Pro Asn Met Asp Arg Glu Ala
        215                 220                 225

AAG GAG GAG TAC CAC GTG GTG ATC CAG GCC AAG GAC ATG GGT GGA CAC     1015
Lys Glu Glu Tyr His Val Val Ile Gln Ala Lys Asp Met Gly Gly His
    230                 235                 240

ATG GGT GGA CTC TCA GGG ACA ACC AAA GTG ACG ATC ACT CTG ACT GAT     1063
Met Gly Gly Leu Ser Gly Thr Thr Lys Val Thr Ile Thr Leu Thr Asp
245                 250                 255                 260

GTC AAC GAC AAC CCA CCA AAG TTT CCA CAG AGC GTG TAC CAG ATG TCT     1111
Val Asn Asp Asn Pro Pro Lys Phe Pro Gln Ser Val Tyr Gln Met Ser
                265                 270                 275

GTA TCA GAA GCA GCT GTC CCG GGG GAG GAA GTA GGA AGG GTG AAG GCT     1159
Val Ser Glu Ala Ala Val Pro Gly Glu Glu Val Gly Arg Val Lys Ala
            280                 285                 290

AAA GAC CCA GAC ATT GGA GAA AAT GGC TTA GTC ACA TAC AAT ATC GTT     1207
Lys Asp Pro Asp Ile Gly Glu Asn Gly Leu Val Thr Tyr Asn Ile Val
        295                 300                 305

GAT GGA GAC GGC ATA GAA CTG TTT GAA ATT ACA ACA GAC TAT GAA ACA     1255
Asp Gly Asp Gly Ile Glu Leu Phe Glu Ile Thr Thr Asp Tyr Glu Thr
    310                 315                 320
```

```
CAG GAT GGT GTG GTG AAG CTG AAA AAG CCT GTA GAT TTT GAA ACC AAA     1303
Gln Asp Gly Val Val Lys Leu Lys Lys Pro Val Asp Phe Glu Thr Lys
325                 330                 335                 340
AGA GCT TAT AGC TTG AAG ATA GAG GCC GCC AAT GTT CAC ATT GAT CCG     1351
Arg Ala Tyr Ser Leu Lys Ile Glu Ala Ala Asn Val His Ile Asp Pro
                345                 350                 355
AAG TTC ATC AGC AAT GGA CCT TTC AAG GAC ACT GTG ACC GTC AAG ATT     1399
Lys Phe Ile Ser Asn Gly Pro Phe Lys Asp Thr Val Thr Val Lys Ile
            360                 365                 370
TCA GTA GAA GAT GCC GAT GAG CCT CCC ATG TTC TTG GCC CCA AGT TAT     1447
Ser Val Glu Asp Ala Asp Glu Pro Pro Met Phe Leu Ala Pro Ser Tyr
        375                 380                 385
ATC CAT GAA GTT CAA GAA AAT GCA GCT GCT GGC ACT GTG GTT GGG AGA     1495
Ile His Glu Val Gln Glu Asn Ala Ala Ala Gly Thr Val Val Gly Arg
    390                 395                 400
GTA CAT GCC AAA GAC CCA GAT GCT GCC AAC AGC CCA ATA AGG TAT TCA     1543
Val His Ala Lys Asp Pro Asp Ala Ala Asn Ser Pro Ile Arg Tyr Ser
405                 410                 415                 420
ATT GAT CGT CAT ACT GAC CTC GAC AGG TTT TTC ACG ATT AAT CCA GAA     1591
Ile Asp Arg His Thr Asp Leu Asp Arg Phe Phe Thr Ile Asn Pro Glu
                425                 430                 435
GAT GGT TTT ATT AAA ACT ACG AAA CCT CTA GAT AGG GAA GAA ACT GCC     1639
Asp Gly Phe Ile Lys Thr Thr Lys Pro Leu Asp Arg Glu Glu Thr Ala
            440                 445                 450
TGG CTC AAC ATC TCT GTC TTC GCA GCA GAA ATT CAC AAC AGA CAT CAG     1687
Trp Leu Asn Ile Ser Val Phe Ala Ala Glu Ile His Asn Arg His Gln
        455                 460                 465
GAA ACC AAA GTC CCA GTG GCC ATC AGG GTC CTG GAT GTC AAT GAC AAT     1735
Glu Thr Lys Val Pro Val Ala Ile Arg Val Leu Asp Val Asn Asp Asn
    470                 475                 480
GCT CCT AAG TTT GCT GCC CCT TAT GAA GGT TTT ATC TGT GAG AGC GAT     1783
Ala Pro Lys Phe Ala Ala Pro Tyr Glu Gly Phe Ile Cys Glu Ser Asp
485                 490                 495                 500
CAC CCC AAG GCA CTC TCC AAC CAG CCA ATA GTT ACA GTT AGT GCA GAT     1831
His Pro Lys Ala Leu Ser Asn Gln Pro Ile Val Thr Val Ser Ala Asp
                505                 510                 515
GAC CAG GAC GAC ACA GCC AAT GGA CCA AGA TTT ATC TTC AGC CTA CCC     1879
Asp Gln Asp Asp Thr Ala Asn Gly Pro Arg Phe Ile Phe Ser Leu Pro
            520                 525                 530
```

```
CCT GAA ATC ATG CAC AAC CCA AAC TTC ACA GTA AGA GAC AAC AGA GAT     1927
Pro Glu Ile Met His Asn Pro Asn Phe Thr Val Arg Asp Asn Arg Asp
        535                 540                 545

AAC ACT GCA GGA GTA TAT GCC CGA CGT GGA GGG TTC AGT CGG CAG AAG     1975
Asn Thr Ala Gly Val Tyr Ala Arg Arg Gly Gly Phe Ser Arg Gln Lys
    550                 555                 560

CAG GAC TTC TAC CTC CTG CCC ATT GTG ATC AGT GAT GGT GGC ATT CCA     2023
Gln Asp Phe Tyr Leu Leu Pro Ile Val Ile Ser Asp Gly Gly Ile Pro
565                 570                 575                 580

CCT ATG AGT AGC ACC AAT ACC CTC ACT ATC AAA GTC TGT GGC TGT GAT     2071
Pro Met Ser Ser Thr Asn Thr Leu Thr Ile Lys Val Cys Gly Cys Asp
                585                 590                 595

GTG AAT GGG GCA CTG TTG TCC TGT AAC GCT GAA GCC TAC ATC CTG AAT     2119
Val Asn Gly Ala Leu Leu Ser Cys Asn Ala Glu Ala Tyr Ile Leu Asn
            600                 605                 610

GCC GGT CTG AGC ACT GGG GCA CTG ATC GCC ATC CTT GCC TGC ATC GTC     2167
Ala Gly Leu Ser Thr Gly Ala Leu Ile Ala Ile Leu Ala Cys Ile Val
        615                 620                 625

ATT CTT CTG GTC ATC GTT GTG CTG TTT GTT ACC CTG AGG AGG CAA AAG     2215
Ile Leu Leu Val Ile Val Val Leu Phe Val Thr Leu Arg Arg Gln Lys
    630                 635                 640

AAA GAA CCA CTC ATT GTA TTT GAA GAG GAG GAT GTC CGT GAG AAC ATC     2263
Lys Glu Pro Leu Ile Val Phe Glu Glu Glu Asp Val Arg Glu Asn Ile
645                 650                 655                 660

ATA ACC TAT GAT GAT GAA GGG GGT GGT GAG GAA GAC ACT GAA GCC TTC     2311
Ile Thr Tyr Asp Asp Glu Gly Gly Gly Glu Glu Asp Thr Glu Ala Phe
                665                 670                 675

GAC ATA GCC ACC CTG CAG AAT CCT GAC GGC ATC AAT GGA TTT ATC CCT     2359
Asp Ile Ala Thr Leu Gln Asn Pro Asp Gly Ile Asn Gly Phe Ile Pro
            680                 685                 690

CGC AAA GAT ATC AAA CCT GAG TAT CAG TAT ATG CCT AGA CCT GGG CTG     2407
Arg Lys Asp Ile Lys Pro Glu Tyr Gln Tyr Met Pro Arg Pro Gly Leu
        695                 700                 705

CGA CCA GCA CCC AAC AGT GTG GAT GTG GAC GAC TTC ATC AAC ACA AGA     2455
Arg Pro Ala Pro Asn Ser Val Asp Val Asp Asp Phe Ile Asn Thr Arg
    710                 715                 720

ATA CAG GAG GCA GAT AAT GAT CCC ACA GCC CCT CCC TAT GAC TCC ATC     2503
Ile Gln Glu Ala Asp Asn Asp Pro Thr Ala Pro Pro Tyr Asp Ser Ile
725                 730                 735                 740
```

```
CAA ATC TAT GGT TAT GAG GGC CGG GGT TCC GTG GCT GGG TCC CTG AGC      2551
Gln Ile Tyr Gly Tyr Glu Gly Arg Gly Ser Val Ala Gly Ser Leu Ser
                745                 750                 755
TCC TTG GAG TCT GCC ACG ACA GAC TCA GAC CTG GAC TAC GAC TAT CTA      2599
Ser Leu Glu Ser Ala Thr Thr Asp Ser Asp Leu Asp Tyr Asp Tyr Leu
            760                 765                 770
CAG AAC TGG GGA CCT CGT TTT AAG AAA CTG GCA GAC TTG TAT GGC TCC      2647
Gln Asn Trp Gly Pro Arg Phe Lys Lys Leu Ala Asp Leu Tyr Gly Ser
        775                 780                 785
AAA GAC ACT TTT GAT GAT GAC TCT TAACAATAAT GTTACAAATT TGGCCTTAAG     2701
Lys Asp Thr Phe Asp Asp Asp Ser
    790                 795
AACTGTGTCT GGCATTCTCA AGAATCTAGA AGATGTGTAA ACAGGTATTT TTTTAAATCA    2761
AGGAAAGGCT CATTTAAAAC AAGCAGAGTT TTACAGAGAG GAAACATTTA ATAAAACTGC    2821
AAGGACATCA AAGTGGAAAA TACTGTGAAG TACCTTTTCC CACTTAAAAA GCAAATATTG    2881
AAGTTGTTTA TCAACTTCAG TAGAAAAAAA AAAACCACTT GGCACACAAA ATATTTAAAT    2941
GAAGGAGAAG TCCACGGTGA ACTTACAATG AAGGGAAATC GTCTATGTGT TAAGAACATC    3001
TAAGTCTCTC TTATTTTATT TTTTAATTTG TCAAAGAAGC TTCCACAAAA TTAGAAAGGA    3061
CAACAGTTCT GAGCTGAAAT TTCGCCTTAA ACTATGGACA CTCTATCTGT AGTGCGTTTT    3121
TAAACTTTGA ATATATAATA TCCAGCCAGC TTAAACCCAT ACAATGTATG TACAATACAA    3181
TGTACAATTA TGTCTCTTGA GCATCAATCT TGTTACTGCT GATTCTTGTA AATCTTTTTG    3241
CTTCTACTTT CATCCTAAAC TAATACGTGC CAGATATAAC TGTCTTGTTT CAGTGAGGAG    3301
CACCCTATTT CTATGTCATT TTTAATGTAT CTATTTGTAC AATTTTAAAG TTCTTATTTT    3361
AGTATACATA CAAATATCAG TATTCTGACA TGTACGAAAA TGTTACAGCA TCACACTTAT    3421
ATTTTATGAA CATTGTACTG TTGCTTTAAT ATGAGCTTCA ATATAAGAAG CAACCTTTGA    3481
AATAAAAAAA AGATTCTTTT TTAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA    3541
AAAAAAAAAA AAAAAAAAAA AAAAAAGCGG CCGCGAATTC                          3581
```

SEQ ID NO: 2
SEQUENCE TYPE: nucleic acid
SEQUENCE LENGTH: 3712 base pairs

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULAR TYPE: cDNA to mRNA

ORIGINAL SOURCE
ORGANISM: Homo sapiens
TISSUE TYPE: osteosarcoma

FEATURE:
SEQUENCE Description: SEQ ID NO: 2:

```
GAATTCGGAG ATCTACAGGC GAGGAGAGAT GCCGCGGGGG CCGCTCGCAG CCGCCGCTGA     60

CTTGTGAATG GGACCGGGAC TGGGGCCGGG ACTGACACCG CAGCGCTTGC CCTGCGCCAG    120

GGACTGGCGG CTCGGAGGTT GCGTCCACCC TCAAGGGCCC CAGAAATCAC TGTGTTTTCA    180

GCTCAGCGGC CCTGTGACAT TCCTTCGTGT TGTCATTTGT TGAGTGACCA ATCAGATGGG    240

TGGAGTGTGT TACAGAAATT GGCAGCAAGT ATCCAATGGG TGAAGAAGAA GCTAACTGGG    300

GACGTGGGCA GCCCTGACGT GATGAGCTCA ACCAGCAGAG ACATTCCATC CCAAGAGAGG    360

TCTGCGTGAC GCGTCCGGGA GGCCACCCTC AGCAAGACCA CCGTACAGTT GGTGGAAGGG    420

GTGACAGCTG CATTCTCCTG TGCCTACCAC GTAACCAAAA ATG AAG GAG AAC TAC       475
                                               Met Lys Glu Asn Tyr
                                                1               5

TGT TTA CAA GCC GCC CTG GTG TGC CTG GGC ATG CTG TGC CAC AGC CAT       523
Cys Leu Gln Ala Ala Leu Val Cys Leu Gly Met Leu Cys His Ser His
            10                  15                  20

GCC TTT GCC CCA GAG CGG CGG GGG CAC CTG CGG CCC TCC TTC CAT GGG       571
Ala Phe Ala Pro Glu Arg Arg Gly His Leu Arg Pro Ser Phe His Gly
        25                  30                  35

CAC CAT GAG AAG GGC AAG GAG GGG CAG GTG CTA CAG CGC TCC AAG CGT       619
His His Glu Lys Gly Lys Glu Gly Gln Val Leu Gln Arg Ser Lys Arg
    40                  45                  50

GGC TGG GTC TGG AAC CAG TTC TTC GTG ATA GAG GAG TAC ACC GGG CCT       667
Gly Trp Val Trp Asn Gln Phe Phe Val Ile Glu Glu Tyr Thr Gly Pro
55                  60                  65

GAC CCC GTG CTT GTG GGC AGG CTT CAT TCA GAT ATT GAC TCT GGT GAT       715
Asp Pro Val Leu Val Gly Arg Leu His Ser Asp Ile Asp Ser Gly Asp
70                  75                  80                  85

GGG AAC ATT AAA TAC ATT CTC TCA GGG GAA GGA GCT GGA ACC ATT TTT       763
Gly Asn Ile Lys Tyr Ile Leu Ser Gly Glu Gly Ala Gly Thr Ile Phe
                90                  95                  100

GTG ATT GAT GAC AAA TCA GGG AAC ATT CAT GCC ACC AAG ACG TTG GAT       811
Val Ile Asp Asp Lys Ser Gly Asn Ile His Ala Thr Lys Thr Leu Asp
            105                 110                 115
```

```
CGA GAA GAG AGA GCC CAG TAC ACG TTG ATG GCT CAG GCG GTG GAC AGG      859
Arg Glu Glu Arg Ala Gln Tyr Thr Leu Met Ala Gln Ala Val Asp Arg
        120                 125                 130

GAC ACC AAT CGG CCA CTG GAG CCA CCG TCG GAA TTC ATT GTC AAG GTC      907
Asp Thr Asn Arg Pro Leu Glu Pro Pro Ser Glu Phe Ile Val Lys Val
    135                 140                 145

CAG GAC ATT AAT GAC AAC CCT CCG GAG TTC CTG CAC GAG ACC TAT CAT      955
Gln Asp Ile Asn Asp Asn Pro Pro Glu Phe Leu His Glu Thr Tyr His
150                 155                 160                 165

GCC AAC GTG CCT GAG AGG TCC AAT GTG GGA ACG TCA GTA ATC CAG GTG     1003
Ala Asn Val Pro Glu Arg Ser Asn Val Gly Thr Ser Val Ile Gln Val
                170                 175                 180

ACA GCT TCA GAT GCA GAT GAC CCC ACT TAT GGA AAT AGC GCC AAG TTA     1051
Thr Ala Ser Asp Ala Asp Asp Pro Thr Tyr Gly Asn Ser Ala Lys Leu
            185                 190                 195

GTG TAC AGT ATC CTC GAA GGA CAA CCC TAT TTT TCG GTG GAA GCA CAG     1099
Val Tyr Ser Ile Leu Glu Gly Gln Pro Tyr Phe Ser Val Glu Ala Gln
        200                 205                 210

ACA GGT ATC ATC AGA ACA GCC CTA CCC AAC ATG GAC AGG GAG GCC AAG     1147
Thr Gly Ile Ile Arg Thr Ala Leu Pro Asn Met Asp Arg Glu Ala Lys
    215                 220                 225

GAG GAG TAC CAC GTG GTG ATC CAG GCC AAG GAC ATG GGT GGA CAT ATG     1195
Glu Glu Tyr His Val Val Ile Gln Ala Lys Asp Met Gly Gly His Met
230                 235                 240                 245

GGC GGA CTC TCA GGG ACA ACC AAA GTG ACG ATC ACA CTG ACC GAT GTC     1243
Gly Gly Leu Ser Gly Thr Thr Lys Val Thr Ile Thr Leu Thr Asp Val
                250                 255                 260

AAT GAC AAC CCA CCA AAG TTT CCG CAG AGC GTA TAC CAG ATA TCT GTG     1291
Asn Asp Asn Pro Pro Lys Phe Pro Gln Ser Val Tyr Gln Ile Ser Val
            265                 270                 275

TCA GAA GCA GCC GTC CCT GGG GAG GAA GTA GGA AGA GTG AAA GCT AAA     1339
Ser Glu Ala Ala Val Pro Gly Glu Glu Val Gly Arg Val Lys Ala Lys
        280                 285                 290

GAT CCA GAC ATT GGA GAA AAT GGC TTA GTC ACA TAC AAT ATT GTT GAT     1387
Asp Pro Asp Ile Gly Glu Asn Gly Leu Val Thr Tyr Asn Ile Val Asp
    295                 300                 305

GGA GAT GGT ATG GAA TCG TTT GAA ATC ACA ACG GAC TAT GAA ACA CAG     1435
Gly Asp Gly Met Glu Ser Phe Glu Ile Thr Thr Asp Tyr Glu Thr Gln
310                 315                 320                 325
```

```
GAG GGG GTG ATA AAG CTG AAA AAG CCT GTA GAT TTT GAA ACC AAA AGA     1483
Glu Gly Val Ile Lys Leu Lys Lys Pro Val Asp Phe Glu Thr Lys Arg
                330                 335                 340
GCC TAT AGC TTG AAG GTA GAG GCA GCC AAC GTG CAC ATC GAC CCG AAG     1531
Ala Tyr Ser Leu Lys Val Glu Ala Ala Asn Val His Ile Asp Pro Lys
            345                 350                 355
TTT ATC AGC AAT GGC CCT TTC AAG GAC ACT GTG ACC GTC AAG ATC GCA     1579
Phe Ile Ser Asn Gly Pro Phe Lys Asp Thr Val Thr Val Lys Ile Ala
        360                 365                 370
GTA GAA GAT GCT GAT GAG CCC CCT ATG TTC TTG GCC CCA AGT TAC ATC     1627
Val Glu Asp Ala Asp Glu Pro Pro Met Phe Leu Ala Pro Ser Tyr Ile
    375                 380                 385
CAC GAA GTC CAA GAA AAT GCA GCT GCT GGC ACC GTG GTT GGG AGA GTG     1675
His Glu Val Gln Glu Asn Ala Ala Ala Gly Thr Val Val Gly Arg Val
390                 395                 400                 405
CAT GCC AAA GAC CCT GAT GCT GCC AAC AGC CCG ATA AGG TAT TCC ATC     1723
His Ala Lys Asp Pro Asp Ala Ala Asn Ser Pro Ile Arg Tyr Ser Ile
                410                 415                 420
GAT CGT CAC ACT GAC CTC GAC AGA TTT TTC ACT ATT AAT CCA GAG GAT     1771
Asp Arg His Thr Asp Leu Asp Arg Phe Phe Thr Ile Asn Pro Glu Asp
            425                 430                 435
GGT TTT ATT AAA ACT ACA AAA CCT CTG GAT AGA GAG GAA ACA GCC TGG     1819
Gly Phe Ile Lys Thr Thr Lys Pro Leu Asp Arg Glu Glu Thr Ala Trp
        440                 445                 450
CTC AAC ATC ACT GTC TTT GCA GCA GAA ATC CAC AAT CGG CAT CAG GAA     1867
Leu Asn Ile Thr Val Phe Ala Ala Glu Ile His Asn Arg His Gln Glu
    455                 460                 465
GCC AAA GTC CCA GTG GCC ATT AGG GTC CTT GAT GTC AAC GAT AAT GCT     1915
Ala Lys Val Pro Val Ala Ile Arg Val Leu Asp Val Asn Asp Asn Ala
470                 475                 480                 485
CCC AAG TTT GCT GCC CCT TAT GAA GGT TTC ATC TGT GAG AGT GAT CAG     1963
Pro Lys Phe Ala Ala Pro Tyr Glu Gly Phe Ile Cys Glu Ser Asp Gln
                490                 495                 500
ACC AAG CCA CTT TCC AAC CAG CCA ATT GTT ACA ATT AGT GCA GAT GAC     2011
Thr Lys Pro Leu Ser Asn Gln Pro Ile Val Thr Ile Ser Ala Asp Asp
            505                 510                 515
AAG GAT GAC ACG GCC AAT GGA CCA AGA TTT ATC TTC AGC CTA CCC CCT     2059
Lys Asp Asp Thr Ala Asn Gly Pro Arg Phe Ile Phe Ser Leu Pro Pro
        520                 525                 530
```

```
GAA ATC ATT CAC AAT CCA AAT TTC ACA GTC AGA GAC AAC CGA GAT AAC      2107
Glu Ile Ile His Asn Pro Asn Phe Thr Val Arg Asp Asn Arg Asp Asn
    535                 540                 545

ACA GCA GGC GTG TAC GCC CGG CGT GGA GGG TTC AGT CGG CAG AAG CAG      2155
Thr Ala Gly Val Tyr Ala Arg Arg Gly Gly Phe Ser Arg Gln Lys Gln
550                 555                 560                 565

GAC TTG TAC CTT CTG CCC ATA GTG ATC AGC GAT GGC GGC ATC CCG CCC      2203
Asp Leu Tyr Leu Leu Pro Ile Val Ile Ser Asp Gly Gly Ile Pro Pro
                570                 575                 580

ATG AGT AGC ACC AAC ACC CTC ACC ATC AAA GTC TGC GGG TGC GAC GTG      2251
Met Ser Ser Thr Asn Thr Leu Thr Ile Lys Val Cys Gly Cys Asp Val
            585                 590                 595

AAC GGG GCA CTG CTC TCC TGC AAC GCA GAG GCC TAC ATT CTG AAC GCC      2299
Asn Gly Ala Leu Leu Ser Cys Asn Ala Glu Ala Tyr Ile Leu Asn Ala
        600                 605                 610

GGC CTG AGC ACA GGC GCC CTG ATC GCC ATC CTC GCC TGC ATC GTC ATT      2347
Gly Leu Ser Thr Gly Ala Leu Ile Ala Ile Leu Ala Cys Ile Val Ile
    615                 620                 625

CTC CTG GTC ATT GTA GTA TTG TTT GTG ACC CTG AGA AGG CAA AAG AAA      2395
Leu Leu Val Ile Val Val Leu Phe Val Thr Leu Arg Arg Gln Lys Lys
630                 635                 640                 645

GAA CCA CTC ATT GTC TTT GAG GAA GAA GAT GTC CGT GAG AAC ATC ATT      2443
Glu Pro Leu Ile Val Phe Glu Glu Glu Asp Val Arg Glu Asn Ile Ile
                650                 655                 660

ACT TAT GAT GAT GAA GGG GGT GGG GAA GAA GAC ACA GAA GCC TTT GAT      2491
Thr Tyr Asp Asp Glu Gly Gly Gly Glu Glu Asp Thr Glu Ala Phe Asp
            665                 670                 675

ATT GCC ACC CTC CAG AAT CCT GAT GGT ATC AAT GGA TTT ATC CCC CGC      2539
Ile Ala Thr Leu Gln Asn Pro Asp Gly Ile Asn Gly Phe Ile Pro Arg
        680                 685                 690

AAA GAC ATC AAA CCT GAG TAT CAG TAC ATG CCT AGA CCT GGG CTC CGG      2587
Lys Asp Ile Lys Pro Glu Tyr Gln Tyr Met Pro Arg Pro Gly Leu Arg
    695                 700                 705

CCA GCG CCC AAC AGC GTG GAT GTC GAT GAC TTC ATC AAC ACG AGA ATA      2635
Pro Ala Pro Asn Ser Val Asp Val Asp Asp Phe Ile Asn Thr Arg Ile
710                 715                 720                 725

CAG GAG GCA GAC AAT GAC CCC ACG GCT CCT CCT TAT GAC TCC ATT CAA      2683
Gln Glu Ala Asp Asn Asp Pro Thr Ala Pro Pro Tyr Asp Ser Ile Gln
                730                 735                 740
```

```
ATC TAC GGT TAT GAA GGC AGG GGC TCA GTG GCC GGG TCC CTG AGC TCC      2731
Ile Tyr Gly Tyr Glu Gly Arg Gly Ser Val Ala Gly Ser Leu Ser Ser
            745                 750                 755
CTA GAG TCG GCC ACC ACA GAT TCA GAC TTG GAC TAT GAT TAT CTA CAG      2779
Leu Glu Ser Ala Thr Thr Asp Ser Asp Leu Asp Tyr Asp Tyr Leu Gln
        760                 765                 770
AAC TGG GGA CCT CGT TTT AAG AAA CTA GCA GAT TTG TAT GGT TCC AAA      2827
Asn Trp Gly Pro Arg Phe Lys Lys Leu Ala Asp Leu Tyr Gly Ser Lys
    775                 780                 785
GAC ACT TTT GAT GAC GAT TCT TAA CAATAACGAT ACAAATTTGG CCTTAAGAAC     2881
Asp Thr Phe Asp Asp Asp Ser
790                 795
TGTGTCTGGC GTTCTCAAGA ATCTAGAAGA TGTGTAAACA GGTATTTTTT TAAATCAAGG    2941
AAAGGCTCAT TTAAAACAGG CAAAGTTTTA CAGAGAGGAT ACATTTAATA AAACTGCGAG    3001
GACATCAAAG TGGTAAATAC TGTGAAATAC CTTTTCTCAC AAAAAGGCAA ATATTGAAGT    3061
TGTTTATCAA CTTCGCTAGA AAAAAAAAAC ACTTGGCATA CAAAATATTT AAGTGAAGGA    3121
GAAGTCTAAC GCTGAACTGA CAATGAAGGG AAATTGTTTA TGTGTTATGA ACATCCAAGT    3181
CTTTCTTCTT TTTTAAGTTG TCAAAGAAGC TTCCACAAAA TTAGAAAAGGA CAACAGTTCT   3241
GAGCTGTAAT TTCGCCTTAA ACTCTGGACA CTCTATATGT AGTGCATTTT TAAACTTGAA    3301
ATATATAATA TTCAGCCAGC TTAAACCCAT ACAATGTATG TACAATACAA TGTACAATTA    3361
TGTCTCTTGA GCATCAATCT TGTTACTGCT GATTCTTGTA AATCTTTTTG CTTCTACTTT    3421
CATCTTAAAC TAATACGTGC CAGATATAAC TGTCTTGTTT CAGTGAGAGA CGCCCTATTT    3481
CTATGTCATT TTTAATGTAT CTATTTGTAC AATTTTAAAG TTCTTATTTT AGTATACATA    3541
TAAATATCAG TATTCTGACA TGTAAGAAAA TGTTACGGCA TCACACTTAT ATTTTATGAA    3601
CATTGTACTG TTGCTTTAAT ATGAGCTTCA ATATAAGAAG CAATCTTTGA AATAAAAAAA    3661
GATTTTTTTT TAAAAAAAAG GAGATCTACA GGCCTGTAGA TCTCCGAATT C             3712
```

SEQ ID NO: 3
SEQUENCE TYPE: nucleic acid
SEQUENCE LENGTH: 3914 base pairs

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULAR TYPE: cDNA to mRNA

ORIGINAL SOURCE
ORGANISM: Homo sapiens
TISSUE TYPE: osteosarcoma

FEATURE:
SEQUENCE Description: SEQ ID NO: 3:

```
GAATTCGGAG ATCTACAGGC CCGCGACGCT CCCCTCAGCT GGCGGCGGCC GCGGAGAGAT      60

GCCGCGGGGG CCGCTCGCAG CCGCCGCTGA CTTGTGAATG GGACCGGGAC TGGGGCCGGG     120

ACTGACACCG CAGCGCTTGC CCTGCGCCAG GGACTGGCGG CTCGGAGGTT GCGTCCACCC     180

TCAAGGGCCC CAGAAATCAC TGTGTTTTCA GCTCAGCGGC CCTGTGACAT TCCTTCGTGT     240

TGTCATTTGT TGAGTGACCA ATCAGATGGG TGGAGTGTGT TACAGAAATT GGCAGCAAGT     300

ATCCAATGGG TGAAGAAGAA GCTAACTGGG GACGTGGGCA GCCCTGACGT GATGAGCTCA     360

ACCAGCAGAG ACATTCCATC CCAAGAGAGG TCTGCGTGAC GCGTCCGGGA GGCCACCCTC     420

AGCAAGACCA CCGTACAGTT GGTGGAAGGG GTGACAGCTG CATTCTCCTG TGCCTACCAC     480

GTAACCAAAA ATG AAG GAG AAC TAC TGT TTA CAA GCC GCC CTG GTG TGC         529
           Met Lys Glu Asn Tyr Cys Leu Gln Ala Ala Leu Val Cys
             1               5                  10

CTG GGC ATG CTG TGC CAC AGC CAT GCC TTT GCC CCA GAG CGG CGG GGG       577
Leu Gly Met Leu Cys His Ser His Ala Phe Ala Pro Glu Arg Arg Gly
     15                  20                  25

CAC CTG CGG CCC TCC TTC CAT GGG CAC CAT GAG AAG GGC AAG GAG GGG       625
His Leu Arg Pro Ser Phe His Gly His His Glu Lys Gly Lys Glu Gly
 30                  35                  40                  45

CAG GTG CTA CAG CGC TCC AAG CGT GGC TGG GTC TGG AAC CAG TTC TTC       673
Gln Val Leu Gln Arg Ser Lys Arg Gly Trp Val Trp Asn Gln Phe Phe
                 50                  55                  60

GTG ATA GAG GAG TAC ACC GGG CCT GAC CCC GTG CTT GTG GGC AGG CTT       721
Val Ile Glu Glu Tyr Thr Gly Pro Asp Pro Val Leu Val Gly Arg Leu
            65                  70                  75

CAT TCA GAT ATT GAC TCT GGT GAT GGG AAC ATT AAA TAC ATT CTC TCA       769
His Ser Asp Ile Asp Ser Gly Asp Gly Asn Ile Lys Tyr Ile Leu Ser
        80                  85                  90

GGG GAA GGA GCT GGA ACC ATT TTT GTG ATT GAT GAC AAA TCA GGG AAC       817
Gly Glu Gly Ala Gly Thr Ile Phe Val Ile Asp Asp Lys Ser Gly Asn
    95                  100                 105
```

```
ATT CAT GCC ACC AAG ACG TTG GAT CGA GAA GAG AGA GCC CAG TAC ACG      865
Ile His Ala Thr Lys Thr Leu Asp Arg Glu Glu Arg Ala Gln Tyr Thr
110                 115                 120                 125

TTG ATG GCT CAG GCG GTG GAC AGG GAC ACC AAT CGG CCA CTG GAG CCA      913
Leu Met Ala Gln Ala Val Asp Arg Asp Thr Asn Arg Pro Leu Glu Pro
                130                 135                 140

CCG TCG GAA TTC ATT GTC AAG GTC CAG GAC ATT AAT GAC AAC CCT CCG      961
Pro Ser Glu Phe Ile Val Lys Val Gln Asp Ile Asn Asp Asn Pro Pro
            145                 150                 155

GAG TTC CTG CAC GAG ACC TAT CAT GCC AAC GTG CCT GAG AGG TCC AAT     1009
Glu Phe Leu His Glu Thr Tyr His Ala Asn Val Pro Glu Arg Ser Asn
        160                 165                 170

GTG GGA ACG TCA GTA ATC CAG GTG ACA GCT TCA GAT GCA GAT GAC CCC     1057
Val Gly Thr Ser Val Ile Gln Val Thr Ala Ser Asp Ala Asp Asp Pro
    175                 180                 185

ACT TAT GGA AAT AGC GCC AAG TTA GTG TAC AGT ATC CTC GAA GGA CAA     1105
Thr Tyr Gly Asn Ser Ala Lys Leu Val Tyr Ser Ile Leu Glu Gly Gln
190                 195                 200                 205

CCC TAT TTT TCG GTG GAA GCA CAG ACA GGT ATC ATC AGA ACA GCC CTA     1153
Pro Tyr Phe Ser Val Glu Ala Gln Thr Gly Ile Ile Arg Thr Ala Leu
                210                 215                 220

CCC AAC ATG GAC AGG GAG GCC AAG GAG GAG TAC CAC GTG GTG ATC CAG     1201
Pro Asn Met Asp Arg Glu Ala Lys Glu Glu Tyr His Val Val Ile Gln
            225                 230                 235

GCC AAG GAC ATG GGT GGA CAT ATG GGC GGA CTC TCA GGG ACA ACC AAA     1249
Ala Lys Asp Met Gly Gly His Met Gly Gly Leu Ser Gly Thr Thr Lys
        240                 245                 250

GTG ACG ATC ACA CTG ACC GAT GTC AAT GAC AAC CCA CCA AAG TTT CCG     1297
Val Thr Ile Thr Leu Thr Asp Val Asn Asp Asn Pro Pro Lys Phe Pro
    255                 260                 265

CAG AGC GTA TAC CAG ATA TCT GTG TCA GAA GCA GCC GTC CCT GGG GAG     1345
Gln Ser Val Tyr Gln Ile Ser Val Ser Glu Ala Ala Val Pro Gly Glu
270                 275                 280                 285

GAA GTA GGA AGA GTG AAA GCT AAA GAT CCA GAC ATT GGA GAA AAT GGC     1393
Glu Val Gly Arg Val Lys Ala Lys Asp Pro Asp Ile Gly Glu Asn Gly
                290                 295                 300

TTA GTC ACA TAC AAT ATT GTT GAT GGA GAT GGT ATG GAA TCG TTT GAA     1441
Leu Val Thr Tyr Asn Ile Val Asp Gly Asp Gly Met Glu Ser Phe Glu
            305                 310                 315
```

```
ATC ACA ACG GAC TAT GAA ACA CAG GAG GGG GTG ATA AAG CTG AAA AAG    1489
Ile Thr Thr Asp Tyr Glu Thr Gln Glu Gly Val Ile Lys Leu Lys Lys
        320                 325                 330

CCT GTA GAT TTT GAA ACC AAA AGA GCC TAT AGC TTG AAG GTA GAG GCA    1537
Pro Val Asp Phe Glu Thr Lys Arg Ala Tyr Ser Leu Lys Val Glu Ala
    335                 340                 345

GCC AAC GTG CAC ATC GAC CCG AAG TTT ATC AGC AAT GGC CCT TTC AAG    1585
Ala Asn Val His Ile Asp Pro Lys Phe Ile Ser Asn Gly Pro Phe Lys
350                 355                 360                 365

GAC ACT GTG ACC GTC AAG ATC GCA GTA GAA GAT GCT GAT GAG CCC CCT    1633
Asp Thr Val Thr Val Lys Ile Ala Val Glu Asp Ala Asp Glu Pro Pro
                370                 375                 380

ATG TTC TTG GCC CCA AGT TAC ATC CAC GAA GTC CAA GAA AAT GCA GCT    1681
Met Phe Leu Ala Pro Ser Tyr Ile His Glu Val Gln Glu Asn Ala Ala
            385                 390                 395

GCT GGC ACC GTG GTT GGG AGA GTG CAT GCC AAA GAC CCT GAT GCT GCC    1729
Ala Gly Thr Val Val Gly Arg Val His Ala Lys Asp Pro Asp Ala Ala
        400                 405                 410

AAC AGC CCG ATA AGG TAT TCC ATC GAT CGT CAC ACT GAC CTC GAC AGA    1777
Asn Ser Pro Ile Arg Tyr Ser Ile Asp Arg His Thr Asp Leu Asp Arg
    415                 420                 425

TTT TTC ACT ATT AAT CCA GAG GAT GGT TTT ATT AAA ACT ACA AAA CCT    1825
Phe Phe Thr Ile Asn Pro Glu Asp Gly Phe Ile Lys Thr Thr Lys Pro
430                 435                 440                 445

CTG GAT AGA GAG GAA ACA GCC TGG CTC AAC ATC ACT GTC TTT GCA GCA    1873
Leu Asp Arg Glu Glu Thr Ala Trp Leu Asn Ile Thr Val Phe Ala Ala
                450                 455                 460

GAA ATC CAC AAT CGG CAT CAG GAA GCC AAA GTC CCA GTG GCC ATT AGG    1921
Glu Ile His Asn Arg His Gln Glu Ala Lys Val Pro Val Ala Ile Arg
            465                 470                 475

GTC CTT GAT GTC AAC GAT AAT GCT CCC AAG TTT GCT GCC CCT TAT GAA    1969
Val Leu Asp Val Asn Asp Asn Ala Pro Lys Phe Ala Ala Pro Tyr Glu
        480                 485                 490

GGT TTC ATC TGT GAG AGT GAT CAG ACC AAG CCA CTT TCC AAC CAG CCA    2017
Gly Phe Ile Cys Glu Ser Asp Gln Thr Lys Pro Leu Ser Asn Gln Pro
    495                 500                 505

ATT GTT ACA ATT AGT GCA GAT GAC AAG GAT GAC ACG GCC AAT GGA CCA    2065
Ile Val Thr Ile Ser Ala Asp Asp Lys Asp Asp Thr Ala Asn Gly Pro
510                 515                 520                 525
```

```
AGA TTT ATC TTC AGC CTA CCC CCT GAA ATC ATT CAC AAT CCA AAT TTC      2113
Arg Phe Ile Phe Ser Leu Pro Pro Glu Ile Ile His Asn Pro Asn Phe
                530                 535                 540
ACA GTC AGA GAC AAC CGA GAT AAC ACA GCA GGC GTG TAC GCC CGG CGT      2161
Thr Val Arg Asp Asn Arg Asp Asn Thr Ala Gly Val Tyr Ala Arg Arg
            545                 550                 555
GGA GGG TTC AGT CGG CAG AAG CAG GAC TTG TAC CTT CTG CCC ATA GTG      2209
Gly Gly Phe Ser Arg Gln Lys Gln Asp Leu Tyr Leu Leu Pro Ile Val
        560                 565                 570
ATC AGC GAT GGC GGC ATC CCG CCC ATG AGT AGC ACC AAC ACC CTC ACC      2257
Ile Ser Asp Gly Gly Ile Pro Pro Met Ser Ser Thr Asn Thr Leu Thr
    575                 580                 585
ATC AAA GTC TGC GGG TGC GAC GTG AAC GGG GCA CTG CTC TCC TGC AAC      2305
Ile Lys Val Cys Gly Cys Asp Val Asn Gly Ala Leu Leu Ser Cys Asn
590                 595                 600                 605
GCA GAG GCC TAC ATT CTG AAC GCC GGC CTG AGC ACA GGC GCC CTG ATC      2353
Ala Glu Ala Tyr Ile Leu Asn Ala Gly Leu Ser Thr Gly Ala Leu Ile
                610                 615                 620
GCC ATC CTC GCC TGC ATC GTC ATT CTC CTG GGT TGC CCA AGC TTA ATG      2401
Ala Ile Leu Ala Cys Ile Val Ile Leu Leu Gly Cys Pro Ser Leu Met
            625                 630                 635
GAA CCC CCC TCT CCC AGG GAA GAC ATG AGA TTG CTT TAT CTG GGC TTC      2449
Glu Pro Pro Ser Pro Arg Glu Asp Met Arg Leu Leu Tyr Leu Gly Phe
        640                 645                 650
CAG CTG ATG CTA TTT TCC TAT GTT AAA GTA AAC AGA AGA TTT TGT CTT      2497
Gln Leu Met Leu Phe Ser Tyr Val Lys Val Asn Arg Arg Phe Cys Leu
    655                 660                 665
CTG GGG GTC TTT ATA AAA CTT CCT TTC CTC TAT GTG GTG GCT ACA GAG      2545
Leu Gly Val Phe Ile Lys Leu Pro Phe Leu Tyr Val Val Ala Thr Glu
670                 675                 680                 685
AGT CCA ACC ACA CTT ACG TCA TTG TAGTATTGTT TGTGACCCTG AGAAGGCAAA      2599
Ser Pro Thr Thr Leu Thr Ser Leu
                690

AGAAAGAACC ACTCATTGTC TTTGAGGAAG AAGATGTCCG TGAGAACATC ATTACTTATG    2659
ATGATGAAGG GGGTGGGGAA GAAGACACAG AAGCCTTTGA TATTGCCACC CTCCAGAATC    2719
CTGATGGTAT CAATGGATTT ATCCCCCGCA AAGACATCAA ACCTGAGTAT CAGTACATGC    2779
CTAGACCTGG GCTCCGGCCA GCGCCCAACA GCGTGGATGT CGATGACTTC ATCAACACGA    2839
GAATACAGGA GGCAGACAAT GACCCCACGG CTCCTCCTTA TGACTCCATT CAAATCTACG    2899
GTTATGAAGG CAGGGGCTCA GTGGCCGGGT CCCTGAGCTC CCTAGAGTCG GCCACCACAG    2959
```

```
ATTCAGACTT GGACTATGAT TATCTACAGA ACTGGGGACC TCGTTTTAAG AAACTAGCAG   3019
ATTTGTATGG TTCCAAAGAC ACTTTTGATG ACGATTCTTA ACAATAACGA TACAAATTTG   3079
GCCTTAAGAA CTGTGTCTGG CGTTCTCAAG AATCTAGAAG ATGTGTAAAC AGGTATTTTT   3139
TTAAATCAAG GAAAGGCTCA TTTAAAACAG GCAAAGTTTT ACAGAGAGGA TACATTTAAT   3199
AAAACTGCGA GGACATCAAA GTGGTAAATA CTGTGAAATA CCTTTTCTCA CAAAAAGGCA   3259
AATATTGAAG TTGTTTATCA ACTTCGCTAG AAAAAAAAAA CACTTGGCAT ACAAAATATT   3319
TAAGTGAAGG AGAAGTCTAA CGCTGAACTG ACAATGAAGG GAAATTGTTT ATGTGTTATG   3379
AACATCCAAG TCTTTCTTCT TTTTTAAGTT GTCAAAGAAG CTTCCACAAA ATTAGAAAGG   3439
ACAACAGTTC TGAGCTGTAA TTTCGCCTTA AACTCTGGAC ACTCTATATG TAGTGCATTT   3499
TTAAACTTGA AATATATAAT ATTCAGCCAG CTTAAACCCA TACAATGTAT GTACAATACA   3559
ATGTACAATT ATGTCTCTTG AGCATCAATC TTGTTACTGC TGATTCTTGT AAATCTTTTT   3619
GCTTCTACTT TCATCTTAAA CTAATACGTG CCAGATATAA CTGTCTTGTT TCAGTGAGAG   3679
ACGCCCTATT TCTATGTCAT TTTTAATGTA TCTATTTGTA CAATTTTAAA GTTCTTATTT   3739
TAGTATACAT ATAAATATCA GTATTCTGAC ATGTAAGAAA ATGTTACGGC ATCACACTTA   3799
TATTTTATGA ACATTGTACT GTTGCTTTAA TATGAGCTTC AATATAAGAA GCAATCTTTG   3859
AAATAAAAAA AGATTTTTTT TTCGGAGATC TACAGGCCTG TAGATCTCCG AATTC        3914
```

SEQ ID NO: 4
SEQUENCE TYPE: nucleic acid
SEQUENCE LENGTH: 21 base pairs

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: other nucleic acid

ORIGINAL SOURCE
ORGANISM: none
STRAIN: none

FEATURE: linker DNA with sequence complementary to SEQ ID NO: 5, termed "ATOS-1"
SEQUENCE Description: SEQ ID NO: 4:

CTCTTGCTTG AATTCGGACT A 21

SEQ ID NO: 5
SEQUENCE TYPE: nucleic acid
SEQUENCE LENGTH: 25 base pairs

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: other nucleic acid

ORIGINAL SOURCE
ORGANISM: none
STRAIN: none

FEATURE: linker DNA with sequence complementary to SEQ ID NO: 4, termed "ATOS-2"
SEQUENCE Description: SEQ ID NO: 5:

TAGTCCGAAT TCAAGCAAGA GCACA 25

SEQ ID NO: 6
SEQUENCE TYPE: nucleic acid
SEQUENCE LENGTH: 21 base pairs

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: other nucleic acid

ORIGINAL SOURCE
ORGANISM: none
STRAIN: none

FEATURE: linker DNA with sequence complementary to SEQ ID NO: 7, termed "ATOS-4"
SEQUENCE Description: SEQ ID NO: 6:

CTCTTGCTTA AGCTTGGACT A 21

SEQ ID NO: 7
SEQUENCE TYPE: nucleic acid
SEQUENCE LENGTH: 25 base pairs

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: other nucleic acid

ORIGINAL SOURCE
ORGANISM: none
STRAIN: none

FEATURE: linker DNA with sequence complementary to SEQ ID NO: 6, termed "ATOS-5"
SEQUENCE Description: SEQ ID NO: 7:

TAGTCCAAGC TTAAGCAAGA GCACA 25

SEQ ID NO: 8
SEQUENCE TYPE: amino acid
SEQUENCE LENGTH: 15 amino acids

TOPOLOGY: linear
MOLECULAR TYPE: peptide

ORIGINAL SOURCE
ORGANISM: Mus musculus

FEATURE: OSF-4.1 (antigen peptide)
LOCATION:
segment of mouse OSF-4 from the 101st to the 115th amino acid residue
SEQUENCE Description: SEQ ID NO: 8:

Phe Val Ile Asp Asp Lys Ser Gly Asn Ile His Ala Thr Lys Thr
5 10 15

Claims

1. A protein comprising mOSF-4 having an amino acid sequence at the 25th to 796th positions in Sequence ID No. 1 of the Sequence Listing, hOSF-4-1 having an amino acid sequence at the 25th to 796th positions in Sequence ID No. 2 of the Sequence Table, or hOSF-4-2 having an amino acid sequence at the 25th to 693rd positions in Sequence ID No. 3 of the Sequence Listing; or an analogue of the mOSF-4, hOSF-4-1 or hOSF-4-2; or a fragment of the mOSF-4, hOSF-4-1 or hOSF-4-2.

2. A protein comprising mOSF-4 precursor protein having an amino acid sequence at the 1st to 796th positions, including a signal peptide at the 1st to 24th positions, in Sequence ID No. 1 of the Sequence Listing, hOSF-4-1 precursor protein having an amino acid sequence at the 1st to 796th positions, including a signal peptide at the 1st to 24th positions, in Sequence ID No. 2 of the Sequence Listing, or hOSF-4-2 precursor protein having an amino acid sequence at the 1st to 693rd positions, including a signal peptide at the 1st to 24th positions, in Sequence ID No. 3 of the Sequence Listing; or an analogue of each precursor protein; or a fragment of each precursor protein.

3. DNA or RNA coding for the protein of Claim 1 or 2.

4. DNA or RNA hybridizing under stingent coditions with DNA or RNA according to claim 3.

5. A process for the production of a recombinant mammalian OSF-4 protein according to claim 1, or an analogue thereof, or a fragment thereof, comprising the steps of:
   (a) obtaining a population of cells containing a heterogeneous DNA composed of the following DNA sequences:
      (i) a sequence which can function in the cells to control transcription and translation, and
      (ii) a DNA sequence joined downstream of said controlling sequence to code for said recombinant protein, and
   (b) culturing said population of cells under conditions which permit the production of said recombinant protein.

6. The process of Claim 5, wherein the controlling sequence further contains a DNA coding for a signal peptide for secreting said recombinant protein extracellularly such that said DNA is positioned immediately upstream of said DNA sequence coding for said recombinant protein.

7. The process of Claim 5 or 6, wherein the population of cells is Escherichia coli, or yeast, or mammalian cells.

8. A diagnostic reagent for bone metabolic diseases, containing the whole or a fragment of the DNA or RNA of Claim 3 or 4.

9. A diagnostic reagent for bone metabolic diseases, containing the protein of Claim 1.

10. A polyclonal or monoclonal antibody against the protein of Claim 1.

11. A diagnostic reagent for bone metabolic diseases, containing the antibody of Claim 10.

12. A therapeutic agent for bone metabolic diseases, containing the protein of Claim 1.

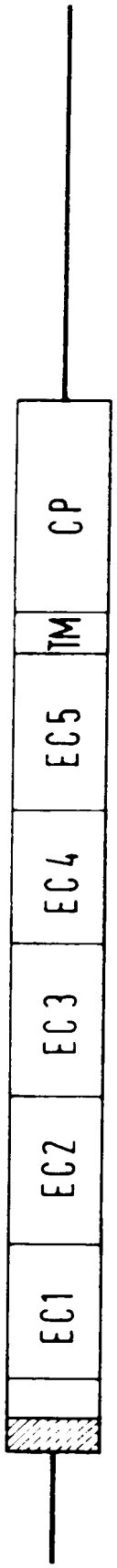


Fig. 1

Fig. 2

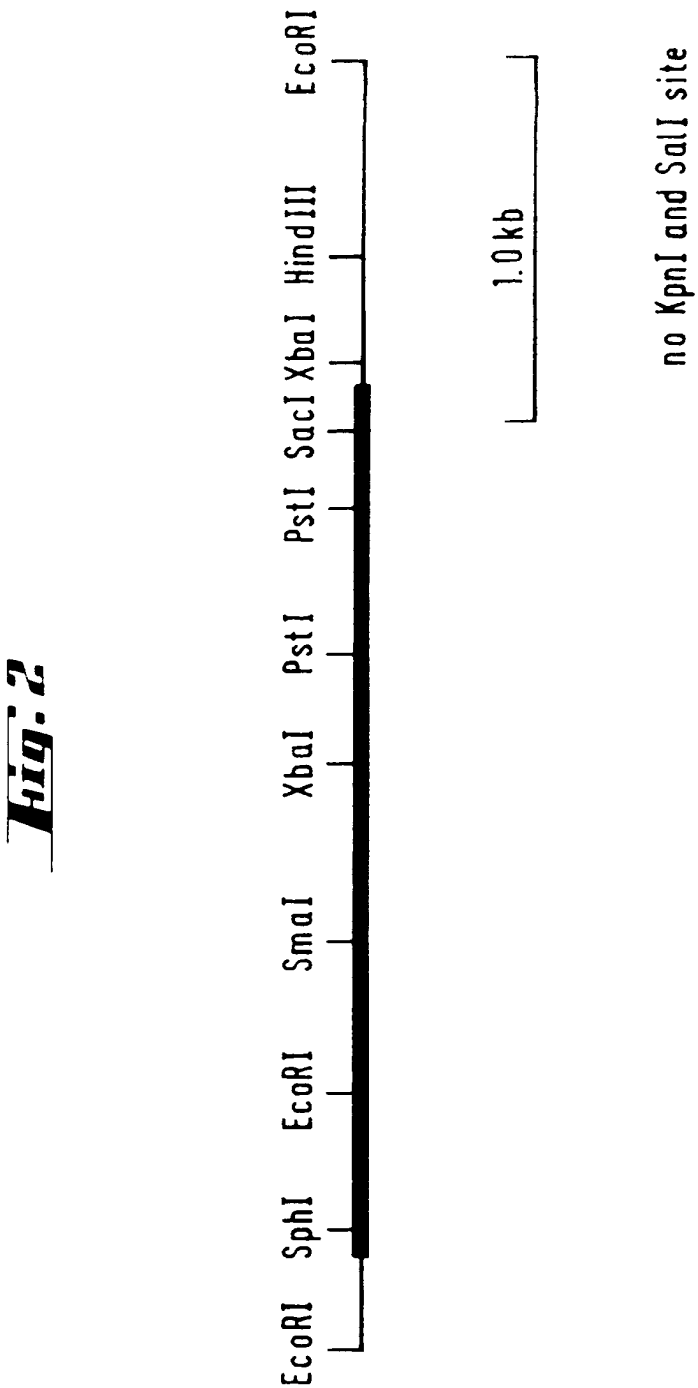
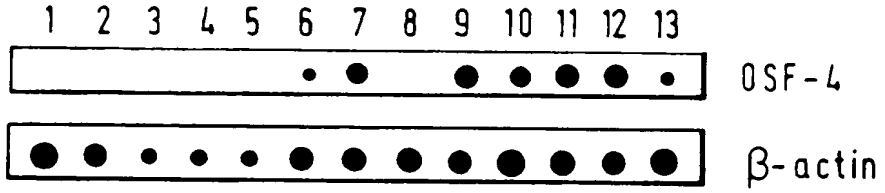

EcoRI
SphI
EcoRI
SmaI
XbaI
PstI
PstI
SacI
XbaI
HindIII
EcoRI
1.0 kb
no KpnI and SalI site

Fig. 3



1 THYMUS

2 SPLEEN

3 BRAIN

4 KIDNEY

5 LIVER

6 LUNG

7 TESTIS

8 HEART

9 OSTEOBLAST-ENRICHED CELL FROM MOUSE CALVARIA

10 MC3T3-E1 CELLS FROM 3 DAYS CULTURE

11 MC3T3-E1 CELLS FROM 12 DAYS CULTURE

12 MC3T3-E1 CELLS FROM 60 DAYS CULTURE

13 NIH3T3 CELLS

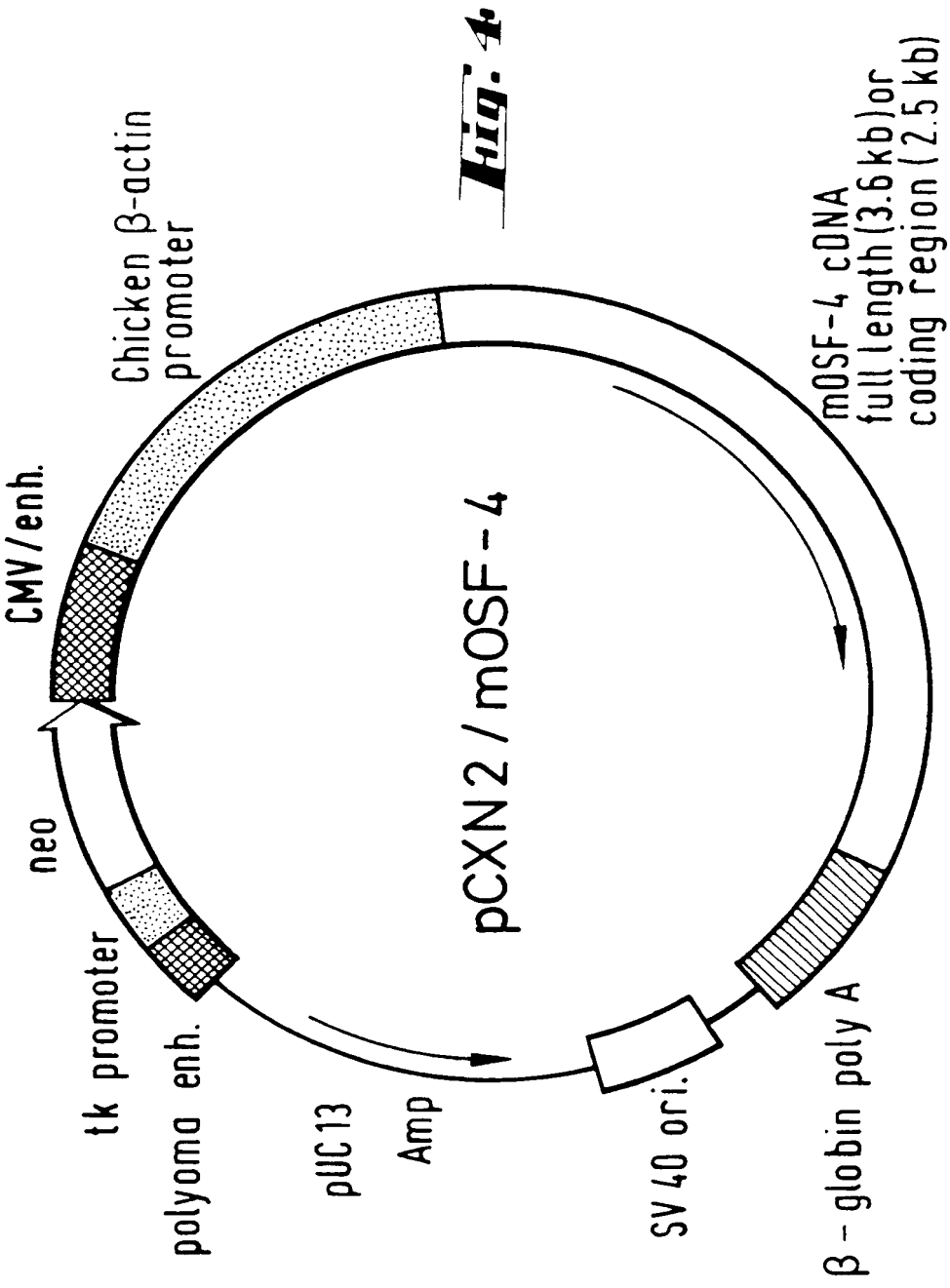

pCXN2 / mOSF-4
CMV/enh.
Chicken β-actin promoter
mOSF-4 cDNA full length(3.6kb)or coding region(2.5kb)
neo
tk promoter
polyoma enh.
pUC13
Amp
SV40 ori.
β-globin poly A

Fig. 4

Table 5: Aggregation Assay of the cloned cell lines expressing mouse OSF-4

Cell lines	No. of cell clots after 1 hr (N1)/ No. of cells pre-stirring (N0)	
	with 1 mM Ca^{2+}	without 1 mM Ca^{2+}
TC-treated		
L-cell (Control)	1.00(*1)	1.00
C1	0.65	1.00
C7	0.29	1.00
C11	0.51	1.00
mock(*2)	1.00	1.00
TE-treated		
C7	1.00	1.00

(*1) The smaller the number of cell clots after one-hour stirring, the stronger cell aggregation occured. The number 1.00 means no aggregation.

(*2) "mock" refers to the control cell into which the vector alone was introduced.

OSF-4 provided by the present invention can be used as an agent for treating bone metabolic diseases, and its high organ specificity for bones enables its use as a diagnostic reagent for bone metabolic diseases.

[Table 1]

```
                1                                                  50                                                 100
mOSF-4      MKENYCLQAA LVCLSMLYHS QAFALERRSH LHPSFHGHHE KGKEGQVLQR SKRGWVWNQF FVIEEYTGPD PVLVGRLHSD IDSGDGNIKY ILSGEGAGTI
hOSF-4-1    MKENYCLQAA LVCLGMLCHS HAFAPERRGH LRPSFHGHHE KGKEGQVLQR SKRGWVWNQF FVIEEYTGPD PVLVGRLHSD IDSGDGNIKY ILSGEGAGTI
hOSF-4-2    MKENYCLQAA LVCLGMLCHS HAFAPERRGH LRPSFHGHHE KGKEGQVLQR SKRGWVWNQF FVIEEYTGPD PVLVGRLHSD IDSGDGNIKY ILSGEGAGTI
Consensus   MKENYCLQAA LVCL-ML-HS -AFA-ERR-H L-PSFHGHHE KGKEGQVLQR SKRGWVWNQF FVIEEYTGPD PVLVGRLHSD IDSGDGNIKY ILSGEGAGTI

                101                                                150                                                200
mOSF-4      FVIDDKSGNI HATKTLDREE RAQYTLMAQA VDRDTNRPLE PPSEFIVKVQ DINDNPPEFL HEIYHANVPE RSNVGTSVIQ VTASDADDPT YGNSAKLVYS
hOSF-4-1    FVIDDKSGNI HATKTLDREE RAQYTLMAQA VDRDTNRPLE PPSEFIVKVQ DINDNPPEFL HETYHANVPE RSNVGTSVIQ VTASDADDPT YGNSAKLVYS
hOSF-4-2    FVIDDKSGNI HATKTLDREE RAQYTLMAQA VDRDTNRPLE PPSEFIVKVQ DINDNPPEFL HETYHANVPE RSNVGTSVIQ VTASDADDPT YGNSAKLVYS
Consensus   FVIDDKSGNI HATKTLDREE RAQYTLMAQA VDRDTNRPLE PPSEFIVKVQ DINDNPPEFL HE-YHANVPE RSNVGTSVIQ VTASDADDPT YGNSAKLVYS

                201                                                250                                                300
mOSF-4      ILEGQPYFSV EAQTGIIRTA LPNMDREAKE EYHVVIQAKD MGGHMGGLSG TTKVTITLTD VNDNPPKFPQ SVYQMSVSEA AVPGEEVGRV KAKDPDIGEN
hOSF-4-1    ILEGQPYFSV EAQTGIIRTA LPNMDREAKE EYHVVIQAKD MGGHMGGLSG TTKVTITLTD VNDNPPKFPQ SVYQISVSEA AVPGEEVGRV KAKDPDIGEN
hOSF-4-2    ILEGQPYFSV EAQTGIIRTA LPNMDREAKE EYHVVIQAKD MGGHMGGLSG TTKVTITLTD VNDNPPKFPQ SVYQISVSEA AVPGEEVGRV KAKDPDIGEN
Consensus   ILEGQPYFSV EAQTGIIRTA LPNMDREAKE EYHVVIQAKD MGGHMGGLSG TTKVTITLTD VNDNPPKFPQ SVYQ-SVSEA AVPGEEVGRV KAKDPDIGEN
```

[Table 2]

```
                301                                                350                                                400
mOSF-4      GLVTYNIVDG DGIELFEITT DYETQDGVVK LKKPVDFETK RAYSLKIEAA NVHIDPKFIS NGPFKDTVTV KISVEDADEP PMFLAPSYIH EVQENAAAGT
hOSF-4-1    GLVTYNIVDG DGMESFEITT DYETQEGVIK LKKPVDFETK RAYSLKVEAA NVHIDPKFIS NGPFKDTVTV KIAVEDADEP PMFLAPSYIH EVQENAAAGT
hOSF-4-2    GLVTYNIVDG DGMESFEITT DYETQEGVIK LKKPVDFETK RAYSLKVEAA NVHIDPKFIS NGPFKDTVTV KIAVEDADEP PMFLAPSYIH EVQENAAAGT
Consensus   GLVTYNIVDG DG-E-FEITT DYETQ-GV-K LKKPVDFETK RAYSLK-EAA NVHIDPKFIS NGPFKDTVTV KI-VEDADEP PMFLAPSYIH EVQENAAAGT

                401                                                450                                                500
mOSF-4      VVGRVHAKDP DAANSPIRYS IDRHTDLDRF FTINPEDGFI KTTKPLDREE TAWLNISVFA AEIHNRHQET KVPVAIRVLD VNDNAPKFAA PYEGFICESD
hOSF-4-1    VVGRVHAKDP DAANSPIRYS IDRHTDLDRF FTINPEDGFI KTTKPLDREE TAWLNITVFA AEIHNRHQEA KVPVAIRVLD VNDNAPKFAA PYEGFICESD
hOSF-4-2    VVGRVHAKDP DAANSPIRYS IDRHTDLDRF FTINPEDGFI KTTKPLDREE TAWLNITVFA AEIHNRHQEA KVPVAIRVLD VNDNAPKFAA PYEGFICESD
Consensus   VVGRVHAKDP DAANSPIRYS IDRHTDLDRF FTINPEDGFI KTTKPLDREE TAWLNI-VFA AEIHNRHQE- KVPVAIRVLD VNDNAPKFAA PYEGFICESD

                501                                                550                                                600
mOSF-4      HPKALSWQPI VTVSADDQDD TANGPRFIFS LPPEIMHNPN FTVRDNRDNT AGVYARRGGF SRQKQDFYLL PIVISDGGIP PMSSTNTLTI KVCGCDVNGA
hOSF-4-1    QTKPLSNQPI VTISADDKDD TANGPRFIFS LPPEIIHNPN FTVRDNRDNT AGVYARRGGF SRQKQDLYLL PIVISDGGIP PMSSTNTLTI KVCGCDVNGA
hOSF-4-2    QTKPLSNQPI VTISADDKDD TANGPRFIFS LPPEIIHNPN FTVRDNRDNT AGVYARRGGF SRQKQDLYLL PIVISDGGIP PMSSTNTLTI KVCGCDVNGA
Consensus   --K-LSNQPI VT-SADD-DD TANGPRFIFS LPPEI-HNPN FTVRDNRDNT AGVYARRGGF SRQKQD-YLL PIVISDGGIP PMSSTNTLTI KVCGCDVNGA
```

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 585 801 A3**

EUROPEAN PATENT APPLICATION

(21) Application number: 93113602.2

(22) Date of filing: 25.08.93

(51) Int. Cl.5 **C12N 15/12**, C07K 13 00, G01N 33 68, C07K 15 28, C12P 21 08, A61K 39 395, G01N 33 577, A61K 37 02

(30) Priority: 28.08.92 JP 230028/92

(43) Date of publication of application: 09.03.94 Bulletin 94/10

(84) Designated Contracting States: AT BE CH DE DK ES FR GB IT LI LU NL PT SE

(88) Date of deferred publication of the search report: 29.06.94 Bulletin 94/26

(71) Applicant: **HOECHST JAPAN LIMITED**
C.P.O. Box 1256
Tokyo 100-91(JP)

(72) Inventor: **Takeshita, Sunao**
1-40-12, Keyakidai
Tokorozawa-shi, Saitama(JP)
Inventor: **Okazaki, Makoto**
Mezon-Ishida 305,
3-12-42 Asahicho
Kawago(JP)
Inventor: **Kawai, Shinji**
1-12-27 Tsurma
Fujimi-shi, Saitama(JP)
Inventor: **Tsujimura, Atsushi**
105 Corpo Haruta,
5 Goshokaido
Mozume(JP)
Inventor: **Amann, Egon, Dr.**
5-1-13 Komazawa
Setagaya-ku, Tokyo(JP)

(74) Representative: **Losert, Wolfgang Dr. et al**
HOECHST AKTIENGESELLSCHAFT,
Zentrale Patentabteilung,
Gebäude F 821
D-65926 Frankfurt am Main (DE)

(54) Bone-related cadherin-like protein and process for its production.

(57) A bone-related protein named OSF-4 which is obtained from bone tissue of a mammal including mouse or human, and a process for its production. This protein is a novel naturally occurring mammal protein of the cadherin family.

OSF-4 acts as an adhesion molecule or a growth factor which takes part in the process of osteogenesis at the site of bone induction. OSF-4 can be used as an agent for treating bone metabolic diseases, and its high organ specificity for bones enables its use as a diagnostic reagent for bone metabolic diseases.

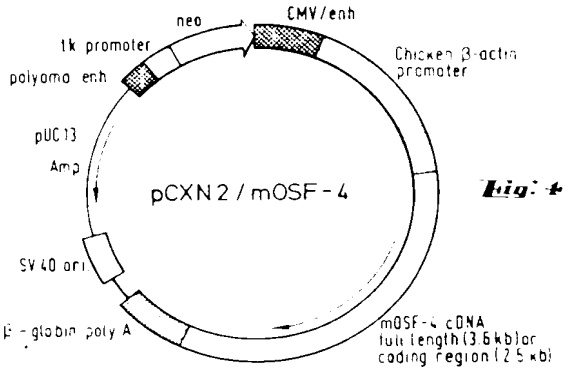


Fig. 4

EP 0 585 801 A3

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 93 11 3602

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.5)
A	BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS. vol. 185, no. 1 , 29 May 1992 , DULUTH, MINNESOTA US pages 224 - 230 'Takamatsu H;Itoh M;Kimura M;Gospodarowicz D;Amann E; "Expression and purification of biologically active human OSF-1 in Escherichia coli"' * the whole document * ---	1-12	C12N15/12 C07K13/00 G01N33/68 C07K15/28 C12P21/08 A61K39/395 G01N33/577 A61K37/02
A	BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS. vol. 173, no. 1 , 30 November 1990 , DULUTH, MINNESOTA US pages 246 - 251 TEZUKA K;TAKESHITA S;HAKEDA Y;KUMEGAWA M;KIKUNO R;HASHIMOTO-GOTOH T; 'Isolation of mouse and human cDNA clones encoding a protein expressed specifically in osteoblasts and brain tissues.' * the whole document * ---	1-12	
A	WO-A-92 00324 (HOECHST JAPAN, LTD.) 9 January 1992 * the whole document * ---	1-12	
X	GENOMICS vol. 12 , 1992 pages 517 - 525 BOYLE, A.L.; WARD, D.C.; 'Isolation and initial characterization of a large repeat sequence element specific to mouse chromosome 8' * figure 8 * --- -/--	1-12	

TECHNICAL FIELDS SEARCHED (Int.Cl.5)

C12N
C07K

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	15 April 1994	Nauche, S

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)